GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 5, 2004, 10:55:48 ; Search time 60 Seconds
(without alignments)
3569.517 Million cell updates/sec

Title: US-09-903-216-2
Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSSGSGS.....IVDVWHPELTPQQRRSLPAI 758

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
1: geneseqp1980s:*
2: geneseqp1990s:*
3: geneseqp2000s:*
4: geneseqp2001s:*
5: geneseqp2002s:*
6: geneseqp2003as:*
7: geneseqp2003bs:*
8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4022	100.0	758	4	AAB83919	Aab83919 A human a
2	4022	100.0	758	6	ABG72365	Abg72365 Human asp
3	4022	100.0	758	6	ADA00639	Ada00639 Human asp
4	1353.5	33.7	265	6	ABU92053	Abu92053 Human pro
5	1334	33.2	255	2	AAY33642	Aay33642 Human lab
6	1320	32.8	255	5	AAU85544	Aau85544 Clone #48
7	1320	32.8	255	6	ABU69516	Abu69516 Human lun
8	1320	32.8	255	6	ABU66419	Abu66419 Lung canc
9	933	23.2	422	4	ABB61986	Abb61986 Drosophil
10	274.5	6.8	369	4	AAB73682	Aab73682 Human oxi
11	271.5	6.8	236	3	AAB43327	Aab43327 Human ORF
12	262	6.5	264	6	ABM68322	Abm68322 Photorhab
13	241	6.0	104	4	AAU29679	Aau29679 Novel hum
14	218.5	5.4	299	6	ABU21099	Abu21099 Protein e
15	217.5	5.4	324	4	AAU28081	Aau28081 Novel hum
16	215	5.3	47	4	AAU31979	Aau31979 Novel hum
17	204.5	5.1	3111	4	ABB60327	Abb60327 Drosophil
18	203.5	5.1	562	2	AAR70491	Aar70491 Leucocyto
19	199	4.9	109	4	AAB73672	Aab73672 Human oxi
20	199	4.9	783	2	AAR05804	Aar05804 C-termina
21	197.5	4.9	1018	2	AAR98747	Aar98747 P. vivax
22	197.5	4.9	1018	2	AAW97039	Aaw97039 A secrete
23	197.5	4.9	1018	4	AAG66528	Aag66528 Plasmodiu
24	197.5	4.9	1132	2	AAR97866	Aar97866 Chicken l
25	196.5	4.9	355	6	ABU23196	Abu23196 Protein e
26	191.5	4.8	1881	3	AAY44506	Aay44506 Streptoco
27	191.5	4.8	1881	6	ABU01047	Abu01047 S. pneumo
28	190.5	4.7	778	3	AAG46504	Aag46504 Arabidops
29	190.5	4.7	788	4	ABB68264	Abb68264 Drosophil
30	190.5	4.7	1162	3	AAY96255	Aay96255 Kaposi's
31	190.5	4.7	1162	3	AAY58500	Aay58500 HHV8 ORF
32	190.5	4.7	1162	4	AAB62331	Aab62331 Amino aci
33	190.5	4.7	1162	5	ABB05621	Abb05621 Kaposi's
34	190	4.7	360	2	AAW03627	Aaw03627 Human fol
35	189.5	4.7	304	6	ABU17264	Abu17264 Protein e
36	189.5	4.7	312	6	ADA36728	Ada36728 Acinetoba
37	189.5	4.7	1616	6	ABU35669	Abu35669 Protein e
38	189.5	4.7	1616	7	ABO23515	Abo23515 Mycoplasm
39	189	4.7	3263	4	ABB67210	Abb67210 Drosophil
40	187.5	4.7	1192	4	ABB59642	Abb59642 Drosophil
41	186.5	4.6	412	2	AAW03626	Aaw03626 Human thy
42	186	4.6	49	6	ADA56732	Ada56732 Human sec
43	186	4.6	49	6	ADA40577	Ada40577 Human sec
44	186	4.6	49	6	ABR47638	Abr47638 Human sec
45	186	4.6	50	2	AAY01148	Aay01148 Secreted

ALIGNMENTS

RESULT 1
AAB83919
ID AAB83919 standard; protein; 758 AA.
XX
AC AAB83919;
XX
DT 23-JUL-2001 (first entry)
XX
DE A human aspartyl (asparaginyl) beta-hydroxylase (HAAH).
XX
KW Epidermal growth factor-like domain; EGF-like domain; cancer;
KW human aspartyl beta-hydroxylase; HAAH; malignant neoplasm; tumour.
XX
OS Homo sapiens.
XX
PN WO200135102-A2.
XX
PD 17-MAY-2001.
XX
PF 08-NOV-2000; 2000WO-US030738.
XX
PR 08-NOV-1999; 99US-00436184.
XX
PA (RHOD-) RHODE ISLAND HOSPITAL LIFESPAN PARTNER.
XX
PI Wands JR, De La Monte SM, Ince N, Carlson RI;
XX
DR WPI; 2001-329171/34.
DR N-PSDB; AAF89811.
XX
PT Diagnosing malignant neoplasm in a mammal, involves contacting mammalian
PT sample with antibody that binds to human aspartyl beta-hydroxylase
PT polypeptide to form antigen-antibody complex and detecting the complex.
XX
PS Disclosure; Page 5; 76pp; English.
XX
CC The present sequence represents a human aspartyl (asparaginyl) beta-
CC hydroxylase (HAAH) enzyme. Epidermal growth factor (EGF)-like domains of
CC polypeptides are hydroxylated by HAAH enzymes. HAAH is used in the method
CC of the invention. The specification describes a method for diagnosing a
CC malignant neoplasm in a mammal. The method comprises contacting a body
CC fluid with an antibody which binds to HAAH polypeptide under complex
CC forming conditions, and detecting the antigen-antibody complex. The
CC method is useful for diagnosing and prognosing a malignant neoplasm in a
CC bodily fluid e.g. central nervous system (CNS)-derived body fluid, blood,
CC serum, urine, saliva, sputum, lung effusion, and ascites fluid of mammal,
CC where the neoplasm is derived from endodermal tissue and is selected from
CC colon cancer, breast cancer, pancreatic cancer, liver cancer, cancer of

```
CC   bile ducts, and cancer of the CNS. HAAH antibodies, linked to a cytotoxic
CC   agent, are useful for killing tumour cells
XX
SQ   Sequence 758 AA;

  Query Match             100.0%;  Score 4022;  DB 4;  Length 758;
  Best Local Similarity   100.0%;  Pred. No. 8.1e-311;
  Matches  758;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||
Db        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758


RESULT 2
ABG72365
ID   ABG72365 standard; protein; 758 AA.
XX
AC   ABG72365;
XX
DT   06-FEB-2003  (first entry)
XX
DE   Human aspartyl (asparaginyl) beta-hydoxylase, HAAH.
XX
KW   Human; enzyme; aspartyl (asparaginyl) beta-hydoxylase; HAAH; cytostatic;
KW   immunostimulant; antibody; neoplasm; tumour; FB50; 86A; 5C7; 19B;
KW   brain tumour; glioma; glioblastoma; astrocytoma; haemangioma;
KW   pancreatic carcinoma; colon cancer; breast cancer; pancreatic cancer;
KW   liver cancer; cancer of the bile ducts; primary malignant CNS neoplasm;
KW   metastatic CNS neoplasm.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Binding-site    286. .291
FT                   /label= Antibody_binding_site
FT                   /note= "This region is specifically claimed in claim 36"
FT   Binding-site    573. .579
FT                   /label= Antibody_binding_site
FT                   /note= "This region is specifically claimed in claim 37"
FT   Domain          600. .700
FT                   /label= Catalytic_domain_1
FT                   /note= "This region is specifically claimed in claim 7"
FT   Binding-site    613. .620
FT                   /label= Antibody_binding_site
FT                   /note= "This region is specifically claimed in claim 38"
FT   Domain          650. .700
FT                   /label= Catalytic_domain_2
FT                   /note= "This region is specifically claimed in claim 24"
FT   Binding-site    660. .758
FT                   /note= "Alpha ketoglutarate binding site. This region is
FT                   specifically claimed in claim 32"
FT   Domain          679. .697
FT                   /note= "Epidermal growth factor-like domain; This region
FT                   is specifically claimed in claim 33"
XX
PN   US2002110559-A1.
XX
PD   15-AUG-2002.
XX
PF   17-MAY-2001; 2001US-00859604.
XX
PR   08-NOV-1999;   99US-00436184.
XX
PA   (WAND/) WANDS J R.
PA   (DMON/) DE LA MONTE S M.
PA   (DEUT/) DEUTCH A H.
PA   (GHAN/) GHANBARI H A.
XX
PI   Wands JR,  De La Monte SM,  Deutch AH,  Ghanbari HA;
XX
DR   WPI; 2003-066676/06.
DR   N-PSDB; ABS57868.
XX
PT   Diagnosing neoplasm and inhibiting tumor growth, by contacting tissue of
PT   mammal with detectably-labeled antibody which binds to human aspartyl
PT   (asparaginyl) beta-hydroxylase.
XX
PS   Claim 14; Page 3; 34pp; English.
XX
CC   The invention relates to diagnosing a neoplasm and inhibiting tumour
CC   growth in a mammal, using an antibody that binds to human aspartyl
CC   (asparaginyl) beta-hydroxylase (HAAH). Diagnosing a neoplasm comprises
CC   contacting a tissue with a detectably-labelled antibody where an increase
CC   in level of antibody binding at tissue site compared to the level of
CC   binding to normal non-neoplastic tissue indicates the presence of a
CC   neoplasm at the tissue site. Inhibiting tumour growth in mammal involves
CC   administering the antibody conjugated to a cytotoxic agent to a mammal.
CC   Also included are a method of conferring an immune response to a tumour
CC   cell in a mammal, by administering the antibody, a method of inducing an
CC   HAAH-specific immune response in a mammal, by administering to the mammal
CC   an HAAH polypeptide (or a polynucleotide composition encoding the
CC   polypeptide, or its degenerate variant), a fragment of HAAH comprising an
CC   extracellular domain and lacking a cytoplasmic domain of HAAH, an
CC   antibody or its fragment which binds to HAAH (where the antibody is FB50,
CC   86A, 5C7 or 19B), a hybridoma cell line chosen from hybridoma FB50,
CC   HA386A, HA15CA and HA219B, and a fragment of HAAH which lacks enzymatic
CC   activity or alpha-ketoglutarate binding domain and epidermal growth
```

CC factor (EGF)-like domain. The methods are useful for diagnosing neoplasm
CC in a mammal, inhibiting tumour growth in a mammal, conferring an immune
CC response to a brain tumour cell (e.g. glioma, glioblastoma, astrocytoma
CC or haemangioma) in a mammal, for conferring immune response to a
CC pancreatic carcinoma cell and for inducing a HAAH-specific immune
CC response in a mammal. The method is useful for diagnosing malignant
CC neoplasms derived from endodermal tissue, e.g. colon cancer, breast
CC cancer, pancreatic cancer, liver cancer and cancer of the bile ducts. The
CC method is also useful for diagnosing neoplasms of central nervous system
CC (CNS) e.g. primary malignant CNS neoplasms of both neuronal and glial
CC cell origin and metastatic CNS neoplasms, and for diagnosing brain
CC tumours e.g. glioma, glioblastoma, astrocytoma or haemangioma. The
CC present sequence represents human HAAH
XX
SQ Sequence 758 AA;

Query Match 100.0%; Score 4022; DB 6; Length 758;
Best Local Similarity 100.0%; Pred. No. 8.1e-311;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||||
Db        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
```

RESULT 3
ADA00639
ID ADA00639 standard; protein; 758 AA.
XX
AC ADA00639;
XX
DT 06-NOV-2003 (first entry)
XX
DE Human aspartyl (asparaginyl) beta-hydroxylase (HAAH).
XX
KW Tumour growth inhibition; human aspartyl (asparaginyl) beta-hydroxylase;
KW HAAH hydroxylation; NOTCH polypeptide;
KW epidermal growth factor-like repeat; EGF-like repeat; tumour cell;
KW malignant neoplasm; colon cancer; breast cancer; pancreatic cancer;
KW liver cancer; cancer of the bile duct; cancer the central nervous system;
KW CNS; cytostatic; enzyme; human.
XX
OS Homo sapiens.
XX
PN US2003031670-A1.
XX
PD 13-FEB-2003.
XX
PF 08-NOV-1999; 99US-00436184.
XX
PR 08-NOV-1999; 99US-00436184.
XX
PA (WAND/) WANDS J R.
PA (DMON/) DE LA MONTE S M.
PA (INCE/) INCE N.
PA (CARL/) CARLSON R I.
XX
PI Wands JR, De La Monte SM, Ince N, Carlson RI;
XX
DR WPI; 2003-605701/57.
DR N-PSDB; ADA00640.
XX
PT Inhibiting tumor growth or killing tumor cells (e.g. cancer of the colon,
PT breast, pancreatic, liver or the central nervous system), by
PT administering an inhibitor of the human aspartyl (asparaginyl) beta-
PT hydroxylase.
XX
PS Disclosure; Page 2; 30pp; English.
XX
CC The present invention relates to a method for inhibiting tumour growth in
CC a mammal. The method comprises administering to the mammal a compound,
CC which inhibits the expression or enzymatic activity of a human aspartyl
CC (asparaginyl) beta-hydroxylase (HAAH). The compound may inhibit HAAH
CC hydroxylation of a NOTCH polypeptide. In particular, the compound may
CC inhibit hydroxylation of an epidermal growth factor (EGF)-like repeat
CC sequence in a NOTCH polypeptide. The methods are useful for inhibiting
CC tumour growth or killing tumour cells, or for diagnosing or
CC prognosticating a malignant neoplasm. In particular, the tumour or
CC neoplasm is colon cancer, breast cancer, pancreatic cancer, liver cancer,
CC cancer of the bile ducts, or cancer or tumour of the central nervous
CC system (CNS). The present sequence represents HAAH.
XX
SQ Sequence 758 AA;

Query Match 100.0%; Score 4022; DB 6; Length 758;
Best Local Similarity 100.0%; Pred. No. 8.1e-311;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||
Db        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
```

RESULT 4
ABU92053
ID ABU92053 standard; protein; 265 AA.
XX
AC ABU92053;
XX
DT 15-JUL-2003 (first entry)
XX
DE Human protein modification and maintenance molecule-33 (PMMM-33).
XX
KW Human; protein modification and maintenance molecule; PMMM; cancer;
KW cell proliferation disorder; atherosclerosis; neurological disorder;
KW epilepsy; Huntington's disease; stroke; immune disorder; allergy;
KW inflammatory disorder; AIDS; developmental disorder; hypothyroidism;
KW Cushing's syndrome; gastrointestinal disorder; epithelial disorder;
KW infection; cytostatic; antiarteriosclerotic; anticonvulsant; nootropic;
KW neuroprotective; cerebroprotective; anti-HIV; antiallergic; vulnerary;
KW antiinflammatory; thyromimetic.
XX
OS Homo sapiens.
XX
PN WO2003031939-A2.
XX
PD 17-APR-2003.
XX
PF 11-OCT-2002; 2002WO-US032850.
XX
PR 12-OCT-2001; 2001US-0329689P.
PR 25-OCT-2001; 2001US-0335703P.
PR 09-NOV-2001; 2001US-0348887P.
PR 28-NOV-2001; 2001US-0334145P.
PR 06-DEC-2001; 2001US-0337451P.
PR 14-DEC-2001; 2001US-0340584P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Ramkumar J, Gorvad AE, Baughn MR, Emerling BM, Yang J, Lee SY;
PI Tran UK, Becha SD, Duggan BM, Lee EA, Griffin JA, Li JX;
PI Sprague WW, Hafalia AJA, Chawla NK, Lehr-Mason PM, Kable AE, Yue H;
PI Marquis JP, Yao MG, Richardson TW, Tang TY, Jin P, Chien D;
PI Bhatia U, Burrill JD, Lee S, Blake JJ, Ho A, Zheng W;
XX
DR WPI; 2003-430274/40.
DR N-PSDB; ACA92448.
XX
PT New human protein modification and maintenance molecules (PMMM), useful
PT for diagnosing, treating and preventing diseases or conditions associated
PT with the aberrant PMMM expression e.g. cancer, atherosclerosis, or
PT infections.
XX
PS Claim 1; Page 273-274; 311pp; English.
XX
CC The present invention relates to the isolation of human protein
CC modification and maintenance molecules (PMMM), and the polynucleotide
CC sequences encoding them. A total of 40 PMMM polypeptides (designated PMMM
CC -1 to PMMM-40) are disclosed. The sequences of the invention are useful
CC for diagnosing a condition or disease associated with the expression of
CC PMMM in a subject, preparing a polyclonal or monoclonal antibody, and
CC generating an expression profile of a sample containing the
CC polynucleotides. The diseases or conditions associated with decreased
CC expression or overexpression of PMMM are cell proliferation disorders
CC (e.g. cancer, atherosclerosis), neurological disorders (e.g. epilepsy,
CC Huntington's disease, stroke), immune/inflammatory disorders, (e.g. AIDS,
CC allergies), developmental disorders (e.g. hypothyroidism, Cushing's
CC syndrome), gastrointestinal or epithelial disorders, and infections. The
CC PMMM polypeptides or their fragments are useful in screening compounds
CC for effectiveness as agonists or antagonists of the polypeptides, or in
CC altering the expression of the target polynucleotide and compounds that
CC specifically bind to, or modulate the activity of the polypeptide.
CC ABU92021-ABU92060 represent the human PMMM polypeptides of the invention
XX
SQ Sequence 265 AA;

Query Match 33.7%; Score 1353.5; DB 6; Length 265;
Best Local Similarity 92.5%; Pred. No. 4.2e-99;
Matches 259; Conservative 2; Mismatches 0; Indels 19; Gaps 1;

```
Qy         34 RETKHGGHKNGRKGGLSGTSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDAD 93
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 KETKHGGHKNGRKGGLSGTSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDAD 64

Qy         94 GDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQS 153
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 GDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQS 124

Qy        154 LLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYH 213
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 LLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYH 184

Qy        214 VEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEG 273
              ||||                          |||||||||||||||||||||||||||||||||
Db        185 VEET------------------DSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEG 225

Qy        274 IEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPET 313
              |||||||||||||||||||||||||||||||||||||:|
Db        226 IEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPDT 265
```

RESULT 5
AAY33642
ID AAY33642 standard; protein; 255 AA.
XX
AC AAY33642;

XX
DT 06-JAN-2000 (first entry)
XX
DE Human labyrinthin protein.
XX
KW Labyrinthin; human; cancer; marker; antigen; detection; antibody;
KW MCA 44-3A6; diagnostic; vaccine; treatment; adenocarcinoma.
XX
OS Homo sapiens.
XX
PN WO9947683-A1.
XX
PD 23-SEP-1999.
XX
PF 11-MAR-1999; 99WO-US005365.
XX
PR 17-MAR-1998; 98US-00040485.
XX
PA (RADO/) RADOSEVICH J A.
XX
PI Radosevich JA;
XX
DR WPI; 1999-580307/49.
DR N-PSDB; AAZ23609.
XX
PT Novel Labyrinthin polynucleotides and polypeptides used as a diagnostic
PT marker for cancer and in anticancer vaccines.
XX
PS Claim 6; Fig 2; 34pp; English.
XX
CC This invention describes a novel polynucleotide and polypeptide sequence
CC for the antigen detected by the antibody MCA 44-3A6. This antigen is
CC designated Labyrinthin (Lab). Antibodies directed to the Labyrinthin
CC (Lab) protein are useful for the detection of Lab. The Lab protein is
CC useful in diagnostic assays for cancer, e.g. to monitor the presence and
CC amount of antibodies (this method is especially useful for cancer cells
CC that have the Lab marker). As the Lab gene is not tissue-specific, it
CC will detect cancer regardless of which organ it occurs in. Peptides
CC derived from Lab are used in the preparation of vaccines to prevent human
CC cancers and/or to treat humans with cancer. Antibody MCA 44-3A6 is able
CC to differentiate antigens associated with adenocarcinomas. However, the
CC sequence of the antigen detected by this antibody has not been elucidated
CC in the prior art. Determination of the polypeptide and polynucleotide
CC sequence of this antigen would enhance its usefulness in cancer
CC diagnosis, treatment and prevention. The present invention discloses the
CC sequences of the antigen recognized by the MCA 44-3A6 antibody. This
CC sequence represents the human lab protein described in the method of the
CC invention
XX
SQ Sequence 255 AA;

Query Match 33.2%; Score 1334; DB 2; Length 255;
Best Local Similarity 99.6%; Pred. No. 1.4e-97;
Matches 254; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```
Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy        179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy        239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 240

Qy        299 SIFPVEEQQEVPPET 313
              |||||||||||||:|
Db        241 SIFPVEEQQEVPPDT 255
```

RESULT 6
AAU85544
ID AAU85544 standard; protein; 255 AA.
XX
AC AAU85544;
XX
DT 21-MAY-2002 (first entry)
XX
DE Clone #48005 (L979P) of lung tumour protein.
XX
KW Lung tumour; cancer; T cell; immune response stimulator; cytostatic.
XX
OS Homo sapiens.
XX
PN WO200204514-A2.
XX
PD 17-JAN-2002.
XX
PF 10-JUL-2001; 2001WO-US022058.
XX
PR 11-JUL-2000; 2000US-00614124.
PR 29-AUG-2000; 2000US-00651563.
PR 08-SEP-2000; 2000US-00658824.
PR 26-SEP-2000; 2000US-00671325.
PR 06-OCT-2000; 2000US-00677419.
PR 30-OCT-2000; 2000US-00702705.
PR 13-DEC-2000; 2000US-00736457.
PR 03-MAY-2001; 2001US-00849626.
XX
PA (CORI-) CORIXA CORP.
XX
PI Wang T, Watanabe Y, Henderson RA, Johnson JC, Retter MW;
PI Marnerakis M, Carter D, Fanger GR, Vedvick TS, Bangur CS, Mcnabb A;
PI Wang A, Fanger N, Switzer A, Mcneill PD, Clapper JD;
XX
DR WPI; 2002-164634/21.
DR N-PSDB; ABK39746.
XX
PT Novel polynucleotide encoding a lung tumor polypeptide useful for
PT stimulating and/or expanding T cells specific for a tumor protein.
XX
PS Claim 2; SEQ ID NO 1806; 223pp; English.
XX
CC The invention describes an isolated polynucleotide and polypeptide useful
CC for stimulating and/or expanding T cells specific for a tumour protein
CC for determining the presence of a cancer in a patient. A composition
CC containing the polynucleotide and/or polypeptide is useful for treating a
CC lung cancer in a patient. The polypeptide is useful for removing tumour
CC cells from a biological sample. The polynucleotide is also useful as
CC probe or primer to detect the level of mRNA encoding a tumour protein.
CC This is the amino acid sequence of a lung tumour associated protein,
CC described in the method of the invention. Note: The sequence data for
CC this patent did not form part of the printed specification, but was
CC obtained in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 255 AA;

Query Match 32.8%; Score 1320; DB 5; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-96;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120
```

```
Qy        179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy        239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy        299 SIFPVEEQQEVPPET 313
              |||||||||||||:|
Db        241 SIFPVEEQQEVPPDT 255
```

RESULT 7
ABU69516
ID ABU69516 standard; protein; 255 AA.
XX
AC ABU69516;
XX
DT 05-JUN-2003 (first entry)
XX
DE Human lung cancer-associated protein L979P.
XX
KW Human; lung cancer; lung tumour; cytostatic; vaccine; T cell expansion;
KW CD4; CD8.
XX
OS Homo sapiens.
XX
PN US2002197669-A1.
XX
PD 26-DEC-2002.
XX
PF 03-MAY-2001; 2001US-00849626.
XX
PR 13-DEC-2000; 2000US-00736457.
XX
PA (BANG/) BANGUR C S.
PA (FANG/) FANGER G R.
PA (WANG/) WANG A.
PA (WANG/) WANG T.
PA (SWIT/) SWITZER A P.
PA (MCNE/) MCNEILL P D.
PA (CLAP/) CLAPPER J D.
XX
PI Bangur CS, Fanger GR, Wang A, Wang T, Switzer AP, Mcneill PD;
PI Clapper JD;
XX
DR WPI; 2003-352750/33.
DR N-PSDB; ACA12072.
XX
PT Novel lung cancer polynucleotide encoding lung cancer protein, useful for
PT detecting the presence of lung cancer in a patient, and in pharmaceutical
PT compositions, e.g. vaccines, for treating lung cancer.
XX
PS Example 5; Page; 72pp; English.
XX
CC The invention relates to a polynucleotide encoding a lung tumour protein,
CC comprising a sequence selected from any of the 14 sequences mentioned in
CC the specification, or a sequence (S2) mentioned in specification,
CC complement of S1, sequences consisting of at least 20 contiguous residues
CC of S1, sequences that hybridise to S1, sequences having 75%, preferably
CC 90%, identity to S1, or degenerate variants of S1. Also included are an
CC isolated polypeptide (comprising a sequence (S3) selected from any one of
CC the 4 amino acid sequences mentioned in the specification, a sequence
CC encoded by the polynucleotide, or sequences having at least 70%,
CC preferably 90%, identity to a sequence encoded by the polynucleotide), an
CC expression vector comprising the polynucleotide operably linked to an
CC expression control sequence, a host cell transformed or transfected with
CC the vector, an isolated antibody (or its antigen-binding fragment) that
CC specifically binds to the polypeptide, detecting the presence of a cancer
CC in a patient, a fusion protein comprising the polypeptide, an
CC oligonucleotide that hybridises to S1 under moderately stringent
CC conditions, stimulating and/or expanding T cells specific for a tumour
CC protein (comprising contacting T cells with the polynucleotide, protein
CC or antigen-presenting cells, under conditions and for a time sufficient
CC to permit the stimulation and/or expansion of T cells) and inhibiting the
CC development of a cancer in a patient (by incubating CD4^+ and/or CD8^+ T
CC cells isolated from a patient with the polynucleotide, protein or antigen
CC presenting cells that express the polynucleotide, such that T cells
CC proliferate, administering to the patient an effective amount of the
CC proliferated T cells, and thus inhibiting the development of a cancer in
CC the patient. The polynucleotide, protein ands cells are useful in a
CC composition for stimulating an immune response in a patient, and for
CC treating a cancer in a patient (particularly lung cancer). The
CC oligonucleotide is useful for determining the presence of a cancer in a
CC patient. The protein and oligonucleotides are useful in pharmaceutical
CC compositions, e.g. vaccines. The polynucleotide is also useful as a probe
CC or primer for nucleic acid hybridisation, and in the design and
CC preparation of ribozyme molecules for inhibiting expression of tumour
CC polypeptides and proteins in tumour cells. An amplified portion of the
CC polynucleotide is useful for isolating a full-length gene from a suitable
CC library. The present sequence is a protein encoded by a cDNA (full
CC length, extended or partial) isolated from a library derived from lung
CC tumour/ cancer cells. Note: The sequence data for this patent did not
CC form part of the printed specification, but was obtained in electronic
CC format directly from the USPTO at
CC seqdata.uspto.gov/sequence.html?DocId=20020197669
XX
SQ Sequence 255 AA;

Query Match 32.8%; Score 1320; DB 6; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-96;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy        179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
              ||||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy        239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db        181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy        299 SIFPVEEQQEVPPET 313
              |||||||||||||:|
Db        241 SIFPVEEQQEVPPDT 255
```

RESULT 8
ABU66419
ID ABU66419 standard; protein; 255 AA.
XX
AC ABU66419;
XX
DT 22-MAY-2003 (first entry)
XX
DE Lung cancer therapyand diagnosis associated protein #43.
XX
KW Lung cancer; cytostatic; vaccine; gene therapy; cancer.
XX
OS Homo sapiens.
XX
PN US2002172952-A1.
XX
PD 21-NOV-2002.
XX
PF 10-JUL-2001; 2001US-00902941.
XX
PR 30-JUN-1999; 99US-00346492.

PR 15-OCT-1999; 99US-00419356.
PR 17-DEC-1999; 99US-00466867.
PR 30-DEC-1999; 99US-00476300.
PR 06-MAR-2000; 2000US-00519642.
PR 22-MAR-2000; 2000US-00533077.
PR 10-APR-2000; 2000US-00546259.
PR 27-APR-2000; 2000US-00560406.
PR 05-JUN-2000; 2000US-00589184.
PR 11-JUL-2000; 2000US-00614124.
PR 29-AUG-2000; 2000US-00651563.
PR 08-SEP-2000; 2000US-00658824.
PR 26-SEP-2000; 2000US-00671325.
PR 06-OCT-2000; 2000US-00677419.
PR 30-OCT-2000; 2000US-00702705.
PR 13-DEC-2000; 2000US-00736457.
PR 03-MAY-2001; 2001US-00849626.
XX
PA (CORI-) CORIXA CORP.
XX
PI Henderson RA, Wang T, Watanabe Y, Johnson JC, Retter MW;
PI Durham M, Carter D, Fanger GR, Vedvick TS, Bangur CS, Mcnabb A;
XX
DR WPI; 2003-328427/31.
XX
PT New polynucleotide, useful for preparing a composition for treating or
PT inhibiting development of cancer, e.g. lung cancer.
XX
PS Example 5; SEQID NO 1807; 82pp; English.
XX
CC The invention describes an isolated polynucleotide comprising one of 32
CC sequences, complement or degenerate variants of them. The polynucleotide
CC is useful for preparing a composition e.g. a vaccine or for gene therapy,
CC for treating or inhibiting development of cancer, e.g. lung cancer. This
CC sequence represents a polypeptide associated with the compositions and
CC methods for the therapy and diagnosis of lung cancer
XX
SQ Sequence 255 AA;

Query Match 32.8%; Score 1320; DB 6; Length 255;
Best Local Similarity 98.8%; Pred. No. 1.9e-96;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy        179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy        239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy        299 SIFPVEEQQEVPPET 313
              |||||||||||||:|
Db        241 SIFPVEEQQEVPPDT 255
```

RESULT 9
ABB61986
ID ABB61986 standard; protein; 422 AA.
XX
AC ABB61986;
XX
DT 26-MAR-2002 (first entry)
XX
DE Drosophila melanogaster polypeptide SEQ ID NO 12750.
XX
KW Drosophila; developmental biology; cell signalling; insecticide;
KW pharmaceutical.
XX
OS Drosophila melanogaster.
XX
PN WO200171042-A2.
XX
PD 27-SEP-2001.
XX
PF 23-MAR-2001; 2001WO-US009231.
XX
PR 23-MAR-2000; 2000US-0191637P.
PR 11-JUL-2000; 2000US-00614150.
XX
PA (PEKE) PE CORP NY.
XX
PI Venter JC, Adams M, Li PWD, Myers EW;
XX
DR WPI; 2001-656860/75.
DR N-PSDB; ABL06089.
XX
PT New isolated nucleic acid detection reagent for detecting 1000 or more
PT genes from Drosophila and for elucidating cell signaling and cell-cell
PT interactions.
XX
PS Disclosure; SEQ ID NO 12750; 21pp + Sequence Listing; English.
XX
CC The invention relates to an isolated nucleic acid detection reagent
CC capable of detecting 1000 or more genes from Drosophila. The invention is
CC useful in developmental biology and in elucidating cell signalling and
CC cell-cell interactions in higher eukaryotes for the development of
CC insecticides, therapeutics and pharmaceutical drugs. The invention
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-
CC ABB72072). The sequence data for this patent did not form part of the
CC printed specification, but was obtained in electronic format directly
CC from WIPO at ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 422 AA;

Query Match 23.2%; Score 933; DB 4; Length 422;
Best Local Similarity 45.0%; Pred. No. 2.5e-65;
Matches 184; Conservative 75; Mismatches 144; Indels 6; Gaps 5;

```
Qy        355 GKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDV 414
              |   :|: :|  |   :   | |  |:|:  : ||:| |||: |  ||: |:   :  ::
Db         15 GNYAQALRSFNTLTTNFAHEPSAHLGRARLLELLAKKERSNQRLWEAIDAYKRYLAFGEL 74

Qy        415 PADLLKLSLKRRS--DRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDN 472
              |   :      |  :  :|||| | :    : |:   | |  |:| | : ||:: :
Db         75 VASNQEFQTAGESCIENLRFLGHHRQATTIHELLINRLPEDPRLRNQLSLTYLMVNNLQQ 134

Qy        473 AKKVYEEVLSVTPNDGFAKVHYGFILKA-QNKIAESIPYLKEGIESGDPGTDDGRFYFHL 531
              :||  | | : ||:  |::|||  |:      |:::||||  :|||: || :  ||  |
Db        135 VEKVAVETLKLWPNNAVAQLHYGLALRQFHADYAKALPYLKYAVESGEEGTQEAFFYLSL 194

Qy        532 GDAMQRVGNK-EAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSL 590
              |: |||:  | || : |  |  :| |||::|||||   |:|||:| ||||||   :: |
Db        195 GETMQRLSMKSEALEVYGKGVAKGFFASLYQRSLYNEPRLRAQPFWQPKETGYERQLEKL 254

Qy        591 ERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTL 650
              ||: |||||||:: :: | |  | | ||:|| | |: |: |||| :: |: || ||:|
Db        255 TLNWRAIRDEGLALLGRS-GFFEDEAELLRDKGVWQQYELYAQGRRVKDNCRRAPITCSL 313

Qy        651 LEKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPK-EGCKIRCANETR 709
              ||:|||: ||||||:|:|:|   |||||| |||||||| || |  |: |   :| | : |
Db        314 LEEFPESAGCRRGQVKFSVMQAKTHVWPHCGPTNCRLRAHLTLAAPEPEKASLRVAEQER 373

Qy        710 TWEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              || ||:: |||||||||||| : |  ||: |:|:|||:|:  |||||  |
Db        374 TWREGELFIFDDSFEHEVWHNGSQSRLVLILDMWHPQLSAAQRRSLSPI 422
```

RESULT 10
AAB73682
ID AAB73682 standard; protein; 369 AA.
XX
AC AAB73682;
XX
DT 11-SEP-2001 (first entry)
XX
DE Human oxidoreductase protein ORP-15.
XX
KW Human oxidoreductase protein; ORP; cell proliferative disorder;
KW arteriosclerosis; cirrhosis; psoriasis; cancer; endocrine disorder;
KW diabetes mellitus; diabetes insipidus; dwarfism; hirsutism; amenorrhoea;
KW osteoporosis; metabolic disorder; obesity; phenylketonuria;
KW hypercholesterolaemia; reproductive disorder; infertility;
KW ovulatory defect; menstrual cycle defect; endometriosis; chromosome 22;
KW polycystic ovary disease; spermatogenesis disruption; impotence;
KW neurological disorder; epilepsy; stroke; Alzheimer's disease;
KW Huntington's disease; Parkinson's disease; Creutzfeldt-Jakob disease;
KW meningitis; cerebral palsy; muscular dystrophy; mood disorder; anxiety;
KW schizophrenic disorder; infection; autoimmune disorder;
KW inflammatory disorder; acquired immunodeficiency syndrome; AIDS; asthma;
KW allergy; Crohn's disease; atopic dermatitis; gout; multiple sclerosis;
KW rheumatoid arthritis; ulcerative colitis; drug screening;
KW toxicity screening; transgenic animal; SNP detection; gene therapy; ss.
XX
OS Homo sapiens.
XX
PN WO200144448-A2.
XX
PD 21-JUN-2001.
XX
PF 07-DEC-2000; 2000WO-US033158.
XX
PR 16-DEC-1999; 99US-0172367P.
XX
PA (INCY-) INCYTE GENOMICS INC.
XX
PI Yue H, Lal P, Tang YT, Hillman JL, Baughn MR, Azimzai Y, Lu DAM;
XX
DR WPI; 2001-390245/41.
DR N-PSDB; AAH24237.
XX
PT Novel human oxidoreductase protein (ORP) useful for diagnosing, treating
PT and preventing cell proliferative, neurological, viral, reproductive and
PT autoimmune/inflammatory disorders associated with abnormal expression of
PT ORP.
XX
PS Claim 1; Page 110-111; 136pp; English.
XX
CC Sequences AAB73668-AAB73694 represent 27 novel human oxidoreductase
CC proteins, designated ORP-1 to ORP-27 respectively, and sequences AAH24223
CC -AAH24249 represent cDNAs encoding ORP-1 to ORP-27. Human ORP proteins
CC and nucleic acids are useful for diagnosing, treating or preventing cell
CC proliferative disorders (e.g. arteriosclerosis, cirrhosis, psoriasis,
CC cancers); endocrine disorders (e.g., type I or II diabetes mellitus,
CC diabetes insipidus, dwarfism, hirsutism, amenorrhoea, osteoporosis);
CC metabolic disorders (e.g., obesity, phenylketonuria,
CC hypercholesterolaemia); reproductive disorders (e.g., infertility,
CC ovulatory and menstrual cycle defects, endometriosis, polycystic ovary
CC disease, disruption of spermatogenesis, impotence); neurological
CC disorders (e.g., epilepsy, stroke, Alzheimer's disease, Huntington's
CC disease, Parkinson's disease, meningitis, Creutzfeldt-Jakob disease,
CC cerebral palsy, muscular dystrophy, mood, anxiety and schizophrenic
CC disorders); viral, bacterial, fungal and parasitic infections; and
CC autoimmune/inflammatory disorders such as acquired immunodeficiency
CC syndrome (AIDS), allergies, asthma, Crohn's disease, atopic dermatitis,
CC gout, multiple sclerosis, rheumatoid arthritis or ulcerative colitis.
CC Human ORP proteins and nucleotides can be used to identify compounds
CC which modulate their activity or expression. ORP nucleic acid sequences
CC may also be used for assessing the toxicity of a test compound, to detect
CC upstream sequences such as promoters and regulatory elements, and to
CC create knock out or knock in animals or transgenic animals to model human
CC disease. Oligonucleotide primers derived from ORP gene sequences may be
CC used to detect single nucleotide polymorphisms (SNPs) and for mapping the
CC naturally occurring genomic sequences. Antibodies specific for ORP
CC proteins may be used in the diagnosis of disorders associated with
CC aberrant ORP expression, in assays to monitor patients being treated with
CC ORP or modulators thereof, and for assessing toxicity of potential drugs
XX
SQ Sequence 369 AA;

Query Match 6.8%; Score 274.5; DB 4; Length 369;
Best Local Similarity 30.1%; Pred. No. 5.4e-13;
Matches 75; Conservative 44; Mismatches 105; Indels 25; Gaps 9;

```
Qy        526 RFYFHLGDAMQRVGNKEAYKWYELG--HK----RGHF----ASVWQRSLYNVNGLKAQPW 575
              : | :| :  :|       | |  :|  ||    :| :     |: :  :: :  |   |:
Db        125 KLYHNLQEYAKR------YSWSGMGRIHKGIREQGRYLNSRPSIQKPEVFFLPDLPTTPY 178

Qy        576 WTPKETGYTELVKSLERNWKLIRDEGLAVMDKAKGLFLPE--DENLREKGDWSQFTLWQQ 633
              ::    :  |: ||||:: |  |  :      ||:    |    |:|  | |  |
Db        179 FSRDAQKHD--VEVLERNFQTILCEFETLYKAFSNCSLPQGWKMNSTPSGEWFTFYLVNQ 236

Qy        634 GRRNENACKGAPKTCTLLEKFPETTGCR-RGQIKYSIMHPGTHVWPHTGPTNCRLRMHLG 692
              |      |:  |:|  ||        |    |    |:: ||| :  | |||| |:| |||
Db        237 GVCVPRNCRKCPRTYRLLGSLRTCIGNNVFGNACISVLSPGTVITEHYGPTNIRIRCHLG 296

Qy        693 LVIPKEGCKIRCANETRTWEEGKVLIFDDSFEHEVWQDASSF---RLIFIVDVWHPELTP 749
              |  |  ||::    | : | ||: |:|||| |  : : |:    |:|:||:||| :
Db        297 LKTP-NGCELVVGGEPQCWAEGRCLLFDDSFLHAAFHEGSAEDGPRVVFMVDLWHPNVAA 355

Qy        750 QQRRSLPAI 758
              :|::|  |
Db        356 AERQALDFI 364
```

RESULT 11
AAB43327
ID AAB43327 standard; protein; 236 AA.
XX
AC AAB43327;
XX
DT 08-FEB-2001 (first entry)
XX
DE Human ORFX ORF3091 polypeptide sequence SEQ ID NO:6182.
XX
KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnerary; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX
OS Homo sapiens.
XX
PN WO200058473-A2.
XX
PD 05-OCT-2000.
XX
PF 31-MAR-2000; 2000WO-US008621.
XX
PR 31-MAR-1999; 99US-0127607P.
PR 02-APR-1999; 99US-0127636P.
PR 05-APR-1999; 99US-0127728P.
PR 30-MAR-2000; 2000US-00540763.
XX
PA (CURA-) CURAGEN CORP.

XX
PI Shimkets RA, Leach M;
XX
DR WPI; 2000-602362/57.
DR N-PSDB; AAC77536.
XX
PT Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease.
XX
PS Claim 11; Page 5365; 5507pp; English.
XX
CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianaemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive
XX
SQ Sequence 236 AA;

Query Match 6.8%; Score 271.5; DB 3; Length 236;
Best Local Similarity 31.2%; Pred. No. 4.7e-13;
Matches 72; Conservative 40; Mismatches 100; Indels 19; Gaps 8;

```
Qy        544 YKWYELG--HK----RGHF----ASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERN 593
              |  |  :|  ||    :| :     |: :  :: :  |   |:::     :   |: ||||
Db          4 YSWSGMGRIHKGIREQGRYLNSRPSIQKPEVFFLPDLPTTPYFSRDAQKHD--VEVLERN  61

Qy        594 WKLIRDEGLAVMDKAKGLFLPE--DENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLL 651
              :: |  |   :          ||:    |    |:|  | |  ||       |:  |:|  ||
Db         62 FQTILCEFETLYKAFSNCSLPQGWKMNSTPSGEWFTFYLVNQGVCVPRNCRKCPRTYRLL 121

Qy        652 EKFPETTGCR-RGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRT 710
                  |    |    |:: ||| :  | |||| |:| ||||  |  ||::    | :
Db        122 GSLRTCIGNNVFGNACISVLSPGTVITEHYGPTNIRIRCHLGLKTP-NGCELVVGGEPQC 180

Qy        711 WEEGKVLIFDDSFEHEVWQDASSF---RLIFIVDVWHPELTPQQRRSLPAI 758
              | ||: |:|||||  |  : : |:    |::|:||:||| :  :|::|  |
Db        181 WAEGRCLLFDDSFLHAAFHEGSAEDGPRVVFMVDLWHPNVAAAERQALDFI 231
```

RESULT 12
ABM68322
ID ABM68322 standard; protein; 264 AA.
XX
AC ABM68322;
XX
DT 20-NOV-2003 (first entry)
XX
DE Photorhabdus luminescens protein sequence #1419.
XX
KW Antibacterial; fungicide; insecticide; polymorphism; genetic analysis;
KW detection; food; gene expression; plant; animal; microorganism; toxin;
KW antibiotic; biopesticide; virulence factor; disease model; plague;
KW whooping cough.
XX
OS Photorhabdus luminescens.
XX
PN WO200294867-A2.
XX
PD 28-NOV-2002.
XX
PF 07-FEB-2002; 2002WO-IB003040.
XX
PR 07-FEB-2001; 2001FR-00001659.
XX
PA (INSP) INST PASTEUR.
PA (CNRS) CNRS CENT NAT RECH SCI.
XX
PI Duchaud E, Taourit S, Glaser P, Frangeul L, Kunst F, Danchin A;
PI Buchrieser C;
XX
DR WPI; 2003-148459/14.
XX
PT Genomic sequence of Photorhabdus luminescens and encoded polypeptides,
PT useful e.g. as therapeutic antimicrobials and agricultural pesticides.
XX
PS Claim 2; SEQ ID NO 1419; 1205pp; French.
XX
CC The invention relates to the isolation of genes and their encoded
CC proteins from Photorhabdus luminescens. The isolated sequences are
CC sources of probes and primers for detecting the genome of P. luminescens
CC and related species; to study polymorphisms; for gene analysis and for
CC detection/amplification of the genes. Antibodies (Ab) raised against the
CC polypeptides encoded by the genes are used for detection/identification
CC of P. luminescens, e.g. in foods. The genes, proteins, Ab and cells that
CC carry a gene-containing vector are used to select compounds that
CC modulate, regulate, induce or inhibit expression of the genes in plants,
CC animals or microorganisms other than P. luminescens and are able to alter
CC response or sensitivity to toxins and antibiotics produced by P.
CC luminescens. Cells transformed to express the genes are useful for
CC recombinant production of the proteins, particularly toxins and
CC antibacterials useful as insecticides, bactericides and fungicides. The
CC genes, proteins, vectors containing the genes and Ab are also useful
CC therapeutically (to treat microbial infection by bacteria or fungi that
CC are sensitive to P. luminescens-encoded toxins or antibiotics) and as
CC biopesticides. Other uses of the genes and the proteins are as virulence
CC factors and for identifying targets of human diseases for which P.
CC luminescens is a model (particularly plague and whooping cough). This
CC sequence represents one of the isolated P. luminescens proteins
XX
SQ Sequence 264 AA;

Query Match 6.5%; Score 262; DB 6; Length 264;
Best Local Similarity 29.1%; Pred. No. 3.2e-12;
Matches 66; Conservative 39; Mismatches 114; Indels 8; Gaps 3;

```
Qy        533 DAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWW-TPKETGYTELVKSLE 591
              | ::|| :     :| |  :|   | :   : | : ||   ||  |         || ||
Db         35 DCLKRVED----MFYPLLQQR---APLQADAKYIMPGLSTTPWLDTNSFPQLQPLVTSLM 87

Qy        592 RNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLL 651
              |   |: |  ||:          :  |  : ||    |:: |: |       | |  :
Db         88 NNADKIKQEINAVISGESQYITDYEHYLGTQKDWKALYLFKNGQPNNAVANILPATWHIF 147

Qy        652 EKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTW 711
                    |   :: :|:: ||| : ||    |  | :|  : ||   |:|  ||| | |
Db        148 NNELRDWHCPLLEVHFSVLQPGTVIKPHCDLWNFTLNLHFAVDIPASHCEIIVANEARCW 207

Qy        712 EEGKVLIFDDSPEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              :||: |:|| |::|| :  :   |:  ::|:||| |:  :| :|  |
Db        208 KEGECLLFDYSYQHEAYNRSDKHRICLLMDIWHPNLSFAEREALVLI 254
```

RESULT 13
AAU29679
ID AAU29679 standard; protein; 104 AA.
XX
AC AAU29679;
XX
DT 18-DEC-2001 (first entry)

XX
DE Novel human secreted protein #170.
XX
KW Human; vaccination; gene therapy; nutritional supplement;
KW stem cell proliferation; haematopoiesis; nerve tissue regeneration;
KW immune suppression; immune stimulation; anti-inflammatory; leukaemia.
XX
OS Homo sapiens.
XX
PN WO200179449-A2.
XX
PD 25-OCT-2001.
XX
PF 16-APR-2001; 2001WO-US008656.
XX
PR 18-APR-2000; 2000US-00552929.
PR 26-JAN-2001; 2001US-00770160.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Drmanac RT;
XX
DR WPI; 2001-611725/70.
XX
PT Nucleic acids encoding a range of human polypeptides, useful in genetic
PT vaccination, testing and therapy.
XX
PS Claim 20; Page 179; 765pp; English.
XX
CC The invention relates to novel human secreted polypeptides. The
CC polypeptides and antibodies to the polypeptides are useful for
CC determining the presence of or predisposition to a disease associated
CC with altered levels of polypeptide. The polypeptides are also useful for
CC identifying agents (agonists and antagonists) that bind to them. Cells
CC expressing the proteins are useful for identifying a therapeutic agent
CC for use in treatment of a pathology related to aberrant expression or
CC physiological interactions of the polypeptide. Vectors comprising the
CC nucleic acids encoding the polypeptides and cells genetically engineered
CC to express them are also useful for producing the proteins. The proteins
CC are useful in genetic vaccination, testing and therapy, and can be used
CC as nutritional supplements. They may be used to increase stem cell
CC proliferation; to regulate haematopoiesis; and in bone, cartilage, tendon
CC and/or nerve tissue growth or regeneration; immune suppression and/or
CC stimulation; as anti-inflammatory agents; and in treatment of leukaemias.
CC AAU29510-AAU33304 represent the amino acid sequences of novel human
CC secreted proteins of the invention
XX
SQ Sequence 104 AA;

Query Match 6.0%; Score 241; DB 4; Length 104;
Best Local Similarity 54.7%; Pred. No. 3.7e-11;
Matches 47; Conservative 11; Mismatches 28; Indels 0; Gaps 0;

Qy 13 SSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMVIALLGVWTSVAV 72
Db 13 SAFTGTCSSTSVRAAAPPGTPQRPSMDAHMTGRKGRLSXTSFFTWSMVTALLGVWTSVSV 72

Qy 73 VWFDLVDYEEVLGKLGIYDADGDGDF 98
Db 73 VWFDLADYDDXLXALAIYDADGDVRF 98

RESULT 14
ABU21099
ID ABU21099 standard; protein; 299 AA.
XX
AC ABU21099;
XX
DT 19-JUN-2003 (first entry)
XX
DE Protein encoded by Prokaryotic essential gene #6626.
XX
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.
XX
OS Burkholderia fungorum.
XX
PN WO200277183-A2.
XX
PD 03-OCT-2002.
XX
PF 21-MAR-2002; 2002WO-US009107.
XX
PR 21-MAR-2001; 2001US-00815242.
PR 06-SEP-2001; 2001US-00948993.
PR 25-OCT-2001; 2001US-0342923P.
PR 08-FEB-2002; 2002US-00072851.
PR 06-MAR-2002; 2002US-0362699P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;
XX
DR WPI; 2003-029926/02.
DR N-PSDB; ACA24969.
XX
PT New antisense nucleic acids, useful for identifying proteins or screening
PT for homologous nucleic acids required for cellular proliferation to
PT isolate candidate molecules for rational drug discovery programs.
XX
PS Claim 25; SEQ ID NO 49023; 1766pp; English.
XX
CC The invention relates to an isolated nucleic acid comprising any one of
CC the 6213 antisense sequences given in the specification where expression
CC of the nucleic acid inhibits proliferation of a cell. Also included are:
CC (1) a vector comprising a promoter operably linked to the nucleic acid
CC encoding a polypeptide whose expression is inhibited by the antisense
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated
CC polypeptide or its fragment whose expression is inhibited by the
CC antisense nucleic acid; (4) an antibody capable of specifically binding
CC the polypeptide;(5) producing the polypeptide; (6) inhibiting cellular
CC proliferation or the activity of a gene in an operon required for
CC proliferation; (7) identifying a compound that influences the activity of
CC the gene product or that has an activity against a biological pathway
CC required for proliferation, or that inhibits cellular proliferation; (8)
CC identifying a gene required for cellular proliferation or the biological
CC pathway in which a proliferation-required gene or its gene product lies
CC or a gene on which the test compound that inhibits proliferation of an
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a
CC compound's activity; (11) a culture comprising strains in which the gene
CC product is overexpressed or underexpressed; (12) determining the extent
CC to which each of the strains is present in a culture or collection of
CC strains; or (13) identifying the target of a compound that inhibits the
CC proliferation of an organism. The antisense nucleic acids are useful for
CC identifying proteins or screening for homologous nucleic acids required
CC for cellular proliferation to isolate candidate molecules for rational
CC drug discovery programs, or for screening homologous nucleic acids
CC required for proliferation in cells other than S. aureus, S. typhimurium,
CC K. pneumoniae or P. aeruginosa. The present sequence is encoded by one of
CC the target prokaryotic essential genes. Note: The sequence data for this
CC patent did not form part of the printed specification, but was obtained
CC in electronic format directly from WIPO at
CC ftp.wipo.int/pub/published_pct_sequences
XX
SQ Sequence 299 AA;

Query Match 5.4%; Score 218.5; DB 6; Length 299;
Best Local Similarity 29.0%; Pred. No. 1.1e-08;
Matches 62; Conservative 29; Mismatches 96; Indels 27; Gaps 6;

Qy 551 HKRGHFASVWQRSL----------YNVNGLKAQPWWTPKETGYTELVKSLERNWKL 596
Db 17 HSRGKVRHGFFRQLSDHSTFTAPLNGFVYLFSALPAQPYLPPSR--FPEL-KLLKEEWRT 73

Qy 597 IRDEGLAVMD----KAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLE 652

```
Db         74 IRDEAFALRDASHIRAATAYNDIGFNSFFRNGWRRFYLKWYGRPHPSAVALCPRTVELLG 133

Qy        653 KFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPK-EGCKIRCANETRTW 711
              : |        |:      ||  :  |  |    || ||||  |  :|| |    || :|
Db        134 RIPSVKAAMFAQLP-----PGGRLGLHRDPYAGALRYHLGLATPNHDGCAIVVDGETYSW 188

Qy        712 EEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHP 745
                :|: ::||::: |  : |    |:|   |:  |
Db        189 RDGEDIVFDETYLHYAFNDTQEDRIILFCDIERP 222
```

RESULT 15
AAU28081
ID AAU28081 standard; protein; 324 AA.
XX
AC AAU28081;
XX
DT 18-DEC-2001 (first entry)
XX
DE Novel human secretory protein, Seq ID No 250.
XX
KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;
KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;
KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;
KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;
KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;
KW gut protection; lung; liver fibrosis; immune deficiency; infection;
KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;
KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;
KW fertility; analgesic; pain; antigen.
XX
OS Homo sapiens.
XX
PN WO200166689-A2.
XX
PD 13-SEP-2001.
XX
PF 05-MAR-2001; 2001WO-US004942.
XX
PR 07-MAR-2000; 2000US-00519705.
PR 19-MAY-2000; 2000US-00574454.
PR 17-JUN-2000; 2000US-00596193.
PR 14-JUL-2000; 2000US-00616847.
PR 19-SEP-2000; 2000US-00665363.
PR 20-OCT-2000; 2000US-00693267.
XX
PA (HYSE-) HYSEQ INC.
XX
PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;
PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;
XX
DR WPI; 2001-589934/66.
DR N-PSDB; AAS44981.
XX
PT Novel polypeptides and nucleic acids obtained from cDNA libraries
PT prepared from various human tissues, for diagnosis and treatment of
PT cancer, neurological, inflammatory, and autoimmune disorders.
XX
PS Example 4; SEQ ID NO 250; 107pp; English.
XX
CC The invention relates to novel isolated human secreted polypeptides (I)
CC and polynucleotides (II). (I) and (II) are useful for treating
CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,
CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is
CC involved in increasing haematopoiesis, stem cell survival, bone growth
CC and remodeling. (I), (II) and modulators of (II) are useful for
CC prophylaxis or treatment of one or more cancers. (II) is also useful for
CC creating transgenic animals useful for studying the in vivo activities of
CC the polypeptide as well as for studying modulators of the polypeptides.
CC (I) induces the proliferation of neural cells and regeneration of nerve
CC and brain tissue and is useful for the treatment of central and
CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,
CC Parkinson's disease, Huntington's disease, and amyotrophic lateral
CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic
CC activity, regulation of haematopoiesis and is useful for treating myeloid
CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve
CC tissue growth, and in tissue repair, healing of burns, incisions, ulcers,
CC for treating osteoporosis, osteoarthritis, bone degenerative disorders,
CC or periodontal disease. Furthermore, (I) is also useful for gut
CC protection or regeneration and treatment of lung or liver fibrosis,
CC reperfusion injury in various tissues, various immune deficiencies and
CC disorders including severe combined immunodeficiency (SCID), bacterial or
CC fungal infections, autoimmune disorders e.g. multiple sclerosis,
CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic
CC reactions and conditions, such as asthma or other respiratory problems.
CC In addition, (I) affects biorhythms or circadian cycles of rhythms,
CC fertility, metabolism, catabolism, anabolism, storage or elimination of
CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides
CC analgesic effects or other pain reducing effects, immunoglobulin like
CC activity and can act as an antigen in a vaccine composition to raise an
CC immune response. AAU28020-AAU28395 represent novel human secreted protein
CC amino acid sequences of the invention
XX
SQ Sequence 324 AA;

Query Match 5.4%; Score 217.5; DB 4; Length 324;
Best Local Similarity 36.6%; Pred. No. 1.5e-08;
Matches 53; Conservative 19; Mismatches 56; Indels 17; Gaps 5;

```
Qy        625 WS--------QFTLWQQGR---RNENACKGAPKTCTLLEKFPETTGCRRGQIKYSIMHPG 673
              ||        |  |:| ||    |  | || :   |  |       |  :|:: ||
Db        175 WSPPLAPGCYQLLLYQAGRCQPSNCRRCPGAYRALRGLRSFMSAN--TFGNAGFSVLLPG 232

Qy        674 THVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFDDSFEHEVWQDAS- 732
                :   |||| |:| |||| ||  ||::    | : | ||  |: |||| | |  : |
Db        233 ARLEGRCGPTNARVRCHLGLKIP-PGCELVVGGEPQCWAEGHCLLVDDSFLHTVAHNGSP 291

Qy        733 --SFRLIFIVDVWHPELTPQQRRSL 755
                |::||||:||| :   :|::|
Db        292 EDGPRVVFIVDLWHPNVAGAERQAL 316
```

Search completed: May 5, 2004, 11:02:20
Job time : 62 secs

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 5, 2004, 11:00:08 ; Search time 21 Seconds
(without alignments)
3472.056 Million cell updates/sec

Title: US-09-903-216-2
Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSSGSGS.....IVDVWHPELTPQQRRSLPAI 758

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_78:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3966.5	98.6	757	2	I38423	aspartyl beta-hydr
2	3172	78.9	754	1	BABOH	peptide-aspartate
3	890	22.1	270	2	JC7792	cardiac junctate-1
4	798	19.8	872	2	T18861	probable peptide-a
5	263.5	6.6	186	2	T47148	hypothetical prote
6	216.5	5.4	1110	2	I51116	NF-180 - sea lampr
7	199	4.9	1271	2	A45555	glutamate rich pro
8	197.5	4.9	312	2	H83527	hypothetical prote
9	191.5	4.8	1881	2	H95076	zinc metalloprotei
10	189.5	4.7	1616	2	G64242	cytadherence-acces
11	188.5	4.7	792	2	T42963	hypothetical prote
12	187	4.6	706	2	A45990	junctional sarcopl
13	183	4.5	763	2	T08929	hypothetical prote
14	181	4.5	771	1	A33430	h-caldesmon - chic
15	179.5	4.5	411	2	S47436	flagellar antigen
16	178.5	4.4	729	2	S68191	triadin - human
17	177	4.4	301	2	A82601	aspartyl/asparagin
18	177	4.4	845	2	A45669	neurofilament trip
19	176	4.4	1094	2	S49313	protein kinase - s
20	174.5	4.3	1948	2	S00485	gene 11-1 protein
21	173.5	4.3	1871	2	D96796	probable heat shoc
22	172.5	4.3	1240	2	S52734	hypothetical prote
23	172	4.3	465	2	S46759	hypothetical prote
24	172	4.3	630	2	S29796	hypothetical prote
25	171.5	4.3	607	2	S27776	80K protein (allel
26	171	4.3	1804	2	T34518	nestin - golden ha
27	169.5	4.2	592	2	B48315	lamin B2 - mouse
28	169.5	4.2	1233	2	S56271	hypothetical prote
29	168.5	4.2	695	2	T40168	hypothetical prote
30	168.5	4.2	913	2	T52485	neurofilament prot
31	168	4.2	302	2	AF1021	probable membrane-
32	168	4.2	1192	2	A71623	probable secreted
33	167.5	4.2	1876	2	E97944	zinc metalloprotei
34	166.5	4.1	1624	2	T25592	hypothetical prote
35	165	4.1	651	2	S18874	nucleolin - Africa
36	164.5	4.1	7962	2	I38346	elastic titin - hu
37	164	4.1	1957	2	T38077	hypothetical coile
38	163.5	4.1	1132	2	T43483	translation initia
39	163	4.1	411	2	S48647	peptidylprolyl iso
40	163	4.1	501	2	C71948	hypothetical prote
41	163	4.1	1877	2	T21861	hypothetical prote
42	163	4.1	1999	1	S21801	myosin heavy chain
43	162.5	4.0	299	2	B83081	hypothetical prote
44	162.5	4.0	793	1	JH0628	caldesmon - human
45	162.5	4.0	992	2	T46337	hypothetical prote

ALIGNMENTS

RESULT 1
I38423
aspartyl beta-hydroxylase - human
C;Species: Homo sapiens (man)
C;Date: 29-May-1998 #sequence_revision 29-May-1998 #text_change 21-Jan-2000
C;Accession: I38423
R;Korioth, F.; Gieffers, C.; Frey, J.
Gene 150, 395-399, 1994
A;Title: Cloning and characterization of the human gene encoding aspartyl beta-hydroxylas
A;Reference number: I38423; MUID:95121937; PMID:7821814
A;Accession: I38423
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-757 <RES>
A;Cross-references: EMBL:U03109; NID:g458031; PIDN:AAA82108.1; PID:g458032
C;Superfamily: peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology
F;54-75/Domain: transmembrane #status predicted <TRM>
F;341-374/Domain: tetratricopeptide repeat homology <TTR>

Query Match 98.6%; Score 3966.5; DB 2; Length 757;
Best Local Similarity 99.2%; Pred. No. 1e-195;
Matches 752; Conservative 2; Mismatches 3; Indels 1; Gaps 1;

```
Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
```

```
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Db        541 KEAYKWYELGHKRGHFASVWQRSLINVNGLKAQP-CGPKETGYTQLVKSLERNWKLIRDE 599

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Db        600 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 659

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Db        660 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETKTWEEGKVLIFD 719

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758

Db        720 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 757
```

RESULT 2
BABOH
peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - bovine
N;Alternate names: aspartyl (asparaginyl) beta-hydroxylase
C;Species: Bos primigenius taurus (cattle)
C;Date: 31-Dec-1993 #sequence_revision 10-Feb-1995 #text_change 11-Jun-1999
C;Accession: A42969; A39470; B39470; C39470; S27948
R;Jia, S.; VanDusen, W.J.; Diehl, R.E.; Kohl, N.E.; Dixon, R.A.; Elliston, K.O.; Stern,
J. Biol. Chem. 267, 14322-14327, 1992
A;Title: cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-hydroxylase.
A;Reference number: A42969; MUID:92332546; PMID:1378441
A;Accession: A42969
A;Molecule type: mRNA
A;Residues: 1-754 <JIA>
A;Cross-references: EMBL:M91213; NID:g162693; PIDN:AAA03563.1; PID:g162694
A;Experimental source: brain
A;Note: sequence extracted from NCBI backbone (NCBIP:108534)
R;Wang, Q.; VanDusen, W.J.; Petroski, C.J.; Garsky, V.M.; Stern, A.M.; Friedman, P.A.
J. Biol. Chem. 266, 14004-14010, 1991
A;Title: Bovine liver aspartyl beta-hydroxylase. Purification and characterization.
A;Reference number: A39470; MUID:91310689; PMID:1856229
A;Accession: A39470
A;Molecule type: protein
A;Residues: 289-328 <WAN>
A;Accession: B39470
A;Molecule type: protein
A;Residues: 615,'X',617-630,'XX',633-634,'X',636,'XX',639-641 <WA2>
A;Accession: C39470
A;Molecule type: protein
A;Residues: 311-347,'X',349,'X',351-373,'X',375-379,'X',381-382 <WA3>
C;Comment: This enzyme uses ferrous iron as a cofactor, and while beta-hydroxylating the
C;Comment: Aspartic acid and asparagine residues in the EGF homology domain of certain p
C;Superfamily: peptide-aspartate beta-dioxygenase; tetratricopeptide repeat homology
C;Keywords: glycoprotein; oxidoreductase; transmembrane protein
F;2-56/Domain: intracellular #status predicted <INC>
F;57-78/Domain: transmembrane #status predicted <TRM>
F;289-754/Product: peptide-aspartate beta-dioxygenase, 56K form #status predicted <56K>
F;311-754/Product: peptide-aspartate beta-dioxygenase, 52K form #status predicted <52K>
F;337-370/Domain: tetratricopeptide repeat homology <TT1>
F;371-404/Domain: tetratricopeptide repeat homology <TT2>
F;13,96,466,702/Binding site: carbohydrate (Asn) (covalent) #status predicted

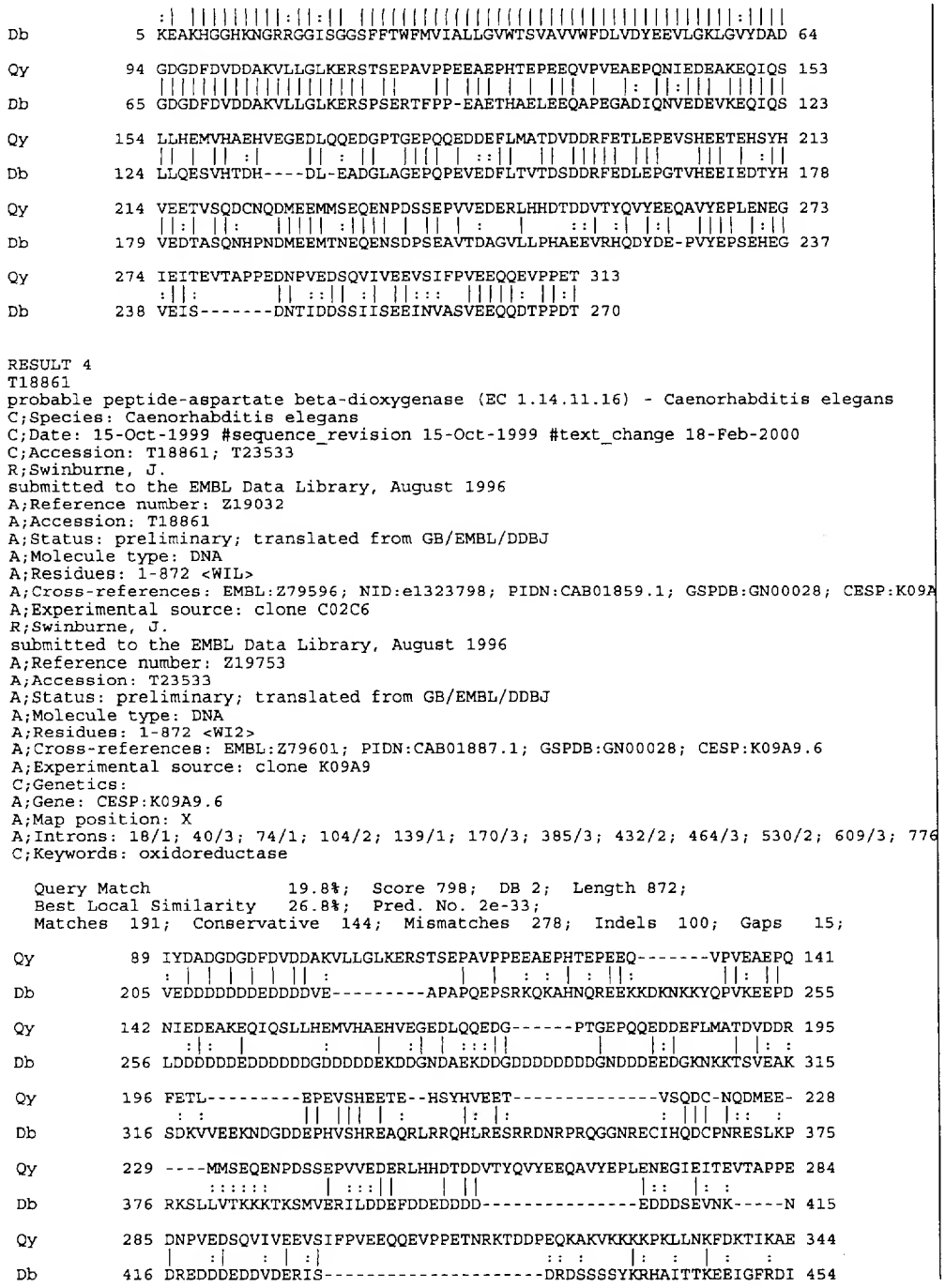
  Query Match 78.9%; Score 3172; DB 1; Length 754;
  Best Local Similarity 78.7%; Pred. No. 4.8e-155;
  Matches 612; Conservative 45; Mismatches 77; Indels 44; Gaps 6;

```
Qy          1 MAQRKNAKSSG---NSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTW 57

Db          1 MAPRKNAKGGGGNSSSSSSGSPTGCTSGGSSSPGARRETKQGGLKNGRKGGLSGSSFFTW 60

Qy         58 FMVIALLGVWTSVAVVWFDLVDYEEVL---------------GKLGIYDADGDGDFDVDD 102

Db         61 FMVIALLGVWTSVAVVWFDLVDYEEVLAKAKDFRYNLSEVLQGKLGIYDADGDGDFDVDD 120

Qy        103 AKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAE 162

Db        121 AKVLLGLKEKPAPKPTVPPEEADMYPWLEDQVLESPGRQNIEDEVYEQVQS-LDETVYSE 179

Qy        163 HVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDC 222

Db        180 --PGENLPQEPEGPAEELQPDDHVFVGSDADDRYEPMGTGAVHEETEDSYHIEETASPAY 237

Qy        223 NQDMEEMMSEQENPDSSEPVVED--ERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVT 280

Db        238 SQDMEDMMYEQENPDSSEPVVVDDAERTYQETDDVTYRDYDEQ---------------- 280

Qy        281 APPEDNPVEDSQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKT 340

Db        281 ----DHAVDNSNTILEEPHMPPAEEQQEVPPETNKKADEPGKKGKVKKKKPKLLNKFDKT 336

Qy        341 IKAELDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRG 400

Db        337 IKAELDAAEKLRKRGKIEEAVNAFEELVRKYPQSPGARYGKAQCEDDLAEKRRSNEILRR 396

Qy        401 AIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDL 460

Db        397 AIETYQEAASLPDAPTDLVKLSLKRRSDRQQFLGHMRGSLLTLQKLVQLFPDDTALKNDL 456

Qy        461 GVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDP 520

Db        457 GVGYLLIGDNDSAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDP 516

Qy        521 GTDDGRFYFHLGDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKE 580

Db        517 GTDDGRFYFHLGDAMQRVGNKEAYRWYELGHQRGHFASVWQRSLYNVQGLKAQPWWTPKE 576

Qy        581 TGYTELVKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENA 640

Db        577 TGYTELVKSLERNWKLIRDEGLAAMDRTHGLFLPEDENLREKGDWSQFTLWQQGRKNENA 636

Qy        641 CKGAPKTCTLLEKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGC 700

Db        637 CKGAPKTCSLLDKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGC 696

Qy        701 KIRCANETRTWEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758

Db        697 KIRCANETRTWEEGKVLIFDDSFEHEVWQDAASFRLIFIVDVWHPELTPHQRRSLPAI 754
```

RESULT 3
JC7792
cardiac junctate-1 - mouse
C;Species: Mus musculus (house mouse)
C;Date: 02-Apr-2002 #sequence_revision 02-Apr-2002 #text_change 03-May-2002
C;Accession: JC7792
R;Hong, C.S.; Kwak, Y.G.; Ji, J.H.; Chae, S.W.; Kim, D.H.
Biochem. Biophys. Res. Commun. 289, 882-887, 2001
A;Title: Molecular cloning and characterization of mouse cardiac junctate isoforms.
A;Reference number: JC7792
A;Contents: Heart
A;Accession: JC7792
A;Molecule type: mRNA
A;Residues: 1-270 <HON>
A;Cross-references: GB:AF302653
C;Comment: This protein, a Ca2+ binding protein. plays a role both in contractile and co
C;Keywords: cardiac muscle; heart

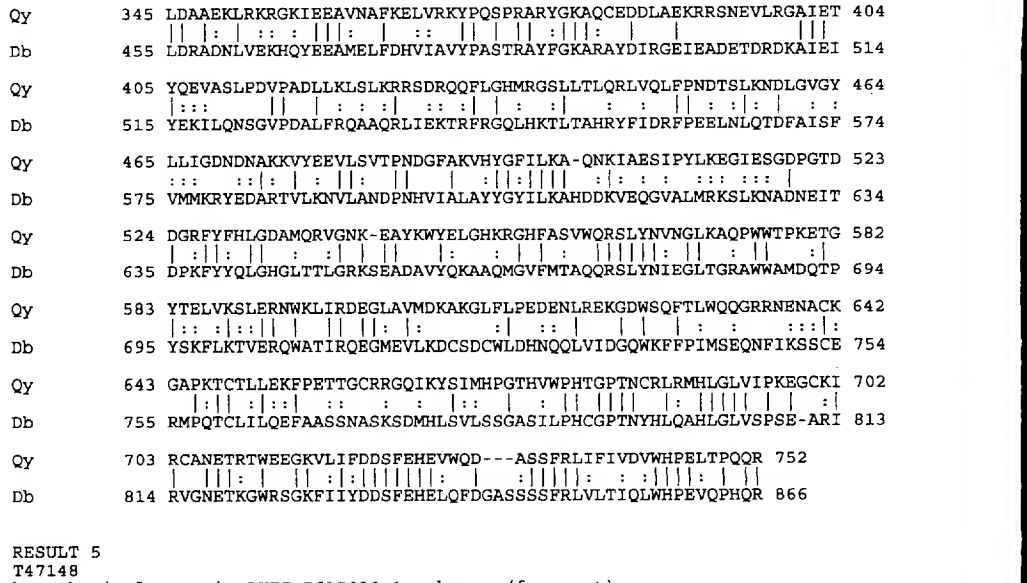
  Query Match 22.1%; Score 890; DB 2; Length 270;
  Best Local Similarity 63.9%; Pred. No. 8.8e-39;
  Matches 179; Conservative 30; Mismatches 57; Indels 14; Gaps 5;

```
Qy         34 RETKHGGHKNGRKGGLSGTSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDAD 93
```



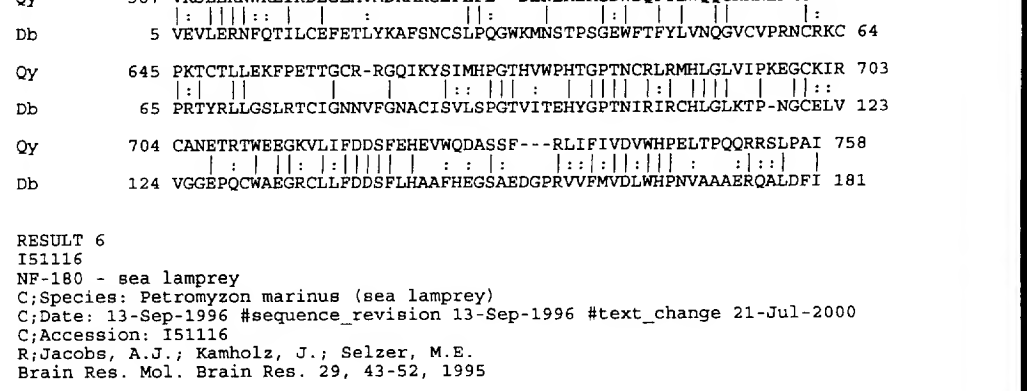
RESULT 4
T18861
probable peptide-aspartate beta-dioxygenase (EC 1.14.11.16) - Caenorhabditis elegans
C;Species: Caenorhabditis elegans
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
C;Accession: T18861; T23533
R;Swinburne, J.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19032
A;Accession: T18861
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-872 <WIL>
A;Cross-references: EMBL:Z79596; NID:e1323798; PIDN:CAB01859.1; GSPDB:GN00028; CESP:K09A
A;Experimental source: clone C02C6
R;Swinburne, J.
submitted to the EMBL Data Library, August 1996
A;Reference number: Z19753
A;Accession: T23533
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-872 <WI2>
A;Cross-references: EMBL:Z79601; PIDN:CAB01887.1; GSPDB:GN00028; CESP:K09A9.6
A;Experimental source: clone K09A9
C;Genetics:
A;Gene: CESP:K09A9.6
A;Map position: X
A;Introns: 18/1; 40/3; 74/1; 104/2; 139/1; 170/3; 385/3; 432/2; 464/3; 530/2; 609/3; 776
C;Keywords: oxidoreductase

Query Match 19.8%; Score 798; DB 2; Length 872;
Best Local Similarity 26.8%; Pred. No. 2e-33;
Matches 191; Conservative 144; Mismatches 278; Indels 100; Gaps 15;



RESULT 5
T47148
hypothetical protein DKFZp761P039.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 20-Apr-2000 #sequence_revision 20-Apr-2000 #text_change 20-Apr-2000
C;Accession: T47148
R;Blum, H.; Bauersachs, S.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A;Reference number: Z24379
A;Accession: T47148
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-186 <AAA>
A;Cross-references: EMBL:AL161993
A;Experimental source: adult amygdala; clone DKFZp761P039
C;Genetics:
A;Note: DKFZp761P039.1

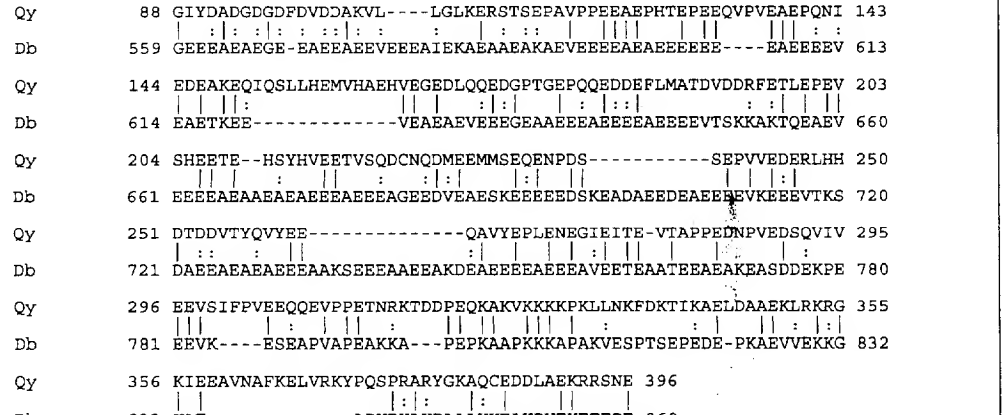
Query Match 6.6%; Score 263.5; DB 2; Length 186;
Best Local Similarity 35.4%; Pred. No. 6.5e-07;
Matches 63; Conservative 28; Mismatches 80; Indels 7; Gaps 4;



RESULT 6
I51116
NF-180 - sea lamprey
C;Species: Petromyzon marinus (sea lamprey)
C;Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C;Accession: I51116
R;Jacobs, A.J.; Kamholz, J.; Selzer, M.E.
Brain Res. Mol. Brain Res. 29, 43-52, 1995

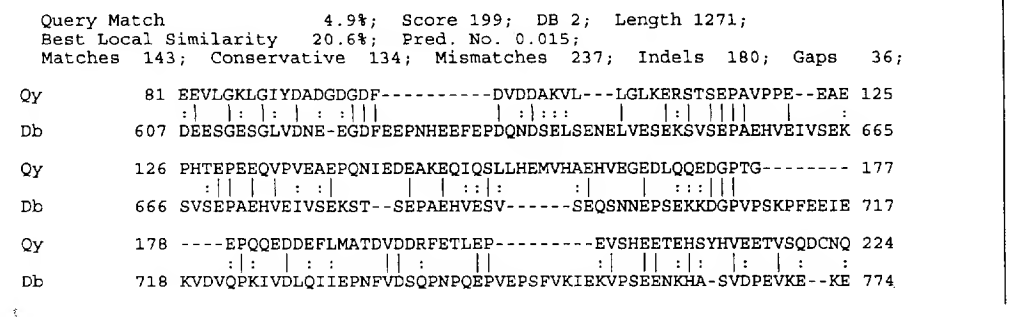
A;Title: The single lamprey neurofilament subunit (NF-180) lacks multiphosphorylation re
A;Reference number: I51116; MUID:95287814; PMID:7770000
A;Accession: I51116
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: mRNA
A;Residues: 1-1110 <JAC>
A;Cross-references: EMBL:U19361; NID:g632548; PIDN:AAA80106.1; PID:g632549
C;Superfamily: neurofilament triplet H protein

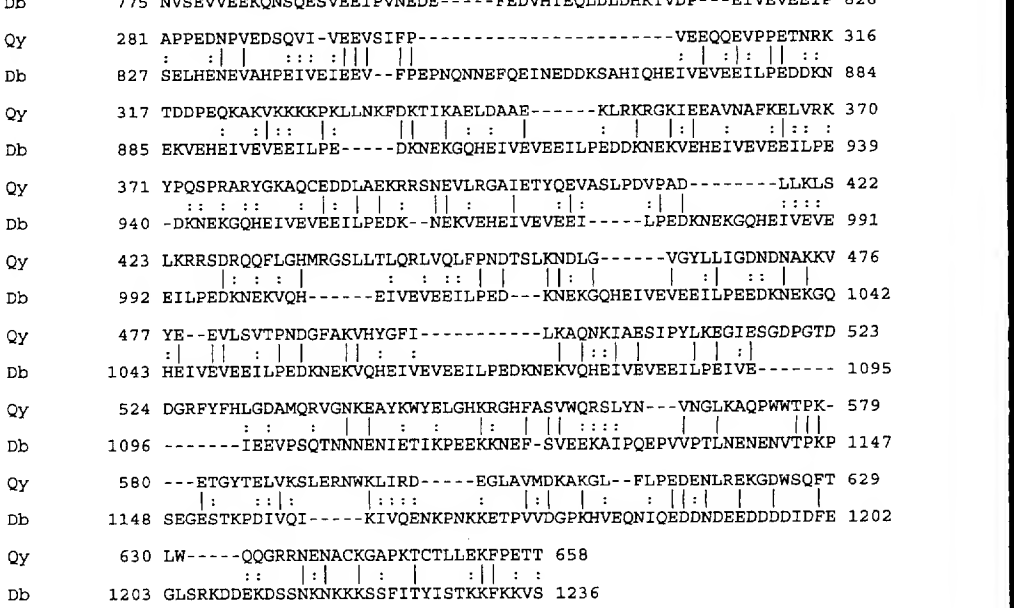
Query Match 5.4%; Score 216.5; DB 2; Length 1110;
Best Local Similarity 26.1%; Pred. No. 0.0016;
Matches 89; Conservative 44; Mismatches 137; Indels 71; Gaps 12;



RESULT 7
A45555
glutamate rich protein - malaria parasite (Plasmodium falciparum)
C;Species: Plasmodium falciparum
C;Date: 17-Feb-1994 #sequence_revision 17-Feb-1994 #text_change 09-Jun-2000
C;Accession: A45555; S27831
R;Borre, M.B.; Dziegiel, M.; Hogh, B.; Petersen, E.; Rieneck, K.; Riley, E.; Meis, J.F.; J.
Mol. Biochem. Parasitol. 49, 119-131, 1991
A;Title: Primary structure and localization of a conserved immunogenic Plasmodium falcip rate life cycle.
A;Reference number: A45555; MUID:92131041; PMID:1775153
A;Accession: A45555
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1271 <BOR>
A;Cross-references: EMBL:M59706; NID:g160311; PID:g160312
A;Note: sequence extracted from NCBI backbone (NCBIN:77801, NCBIP:77802)

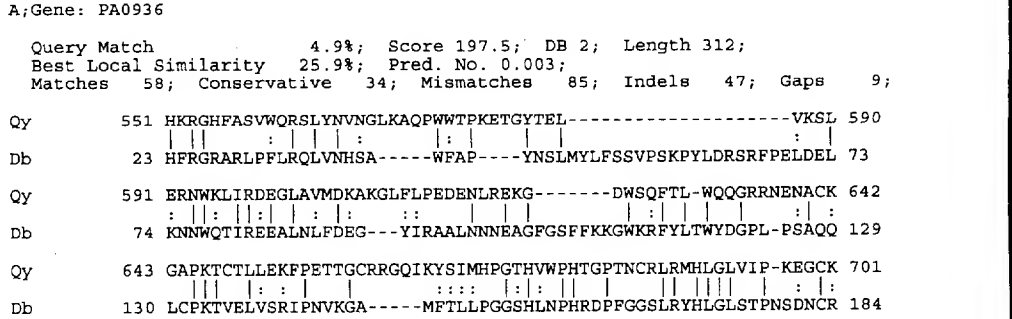
Query Match 4.9%; Score 199; DB 2; Length 1271;
Best Local Similarity 20.6%; Pred. No. 0.015;
Matches 143; Conservative 134; Mismatches 237; Indels 180; Gaps 36;





RESULT 8
H83527
hypothetical protein PA0936 [imported] - Pseudomonas aeruginosa (strain PAO1)
C;Species: Pseudomonas aeruginosa
C;Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C;Accession: H83527
R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrener, P.; Hickey, M.J.; Br adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, .; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A;Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A;Reference number: A82950; MUID:20437337; PMID:10984043
A;Accession: H83527
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-312 <STO>
A;Cross-references: GB:AE004528; GB:AE004091; NID:g9946838; PIDN:AAG04325.1; GSPDB:GN001
A;Experimental source: strain PAO1
C;Genetics:
A;Gene: PA0936

Query Match 4.9%; Score 197.5; DB 2; Length 312;
Best Local Similarity 25.9%; Pred. No. 0.003;
Matches 58; Conservative 34; Mismatches 85; Indels 47; Gaps 9;



```
Qy        702 IRCANETRTWEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHP 745
              |    :   | :|: ::||::| | |  :     |:|   |:  |
Db        185 IYVDGQPYAWRDGEDVMFDETFVHWVKNETEQTRVILFCDIERP 228
```

RESULT 9
H95076
zinc metalloproteinase ZmpB, probable [imported] - Streptococcus pneumoniae (strain TIGR
C;Species: Streptococcus pneumoniae
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: H95076
R;Tettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzapple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A;Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A;Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A;Reference number: A95000; MUID:21357209; PMID:11463916
A;Accession: H95076
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1881 <KUR>
A;Cross-references: GB:AE005672; PIDN:AAK74809.1; PID:g14972138; GSPDB:GN00164; TIGR:SP4
A;Experimental source: strain TIGR4
C;Genetics:
A;Gene: SP0664

Query Match 4.8%; Score 191.5; DB 2; Length 1881;
Best Local Similarity 21.6%; Pred. No. 0.06;
Matches 138; Conservative 82; Mismatches 226; Indels 193; Gaps 30;

```
Qy        111 ERSTSEPAVPPEE--AEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGED 168
              :  ::||| | ||   |  ::|||:| |: | |  :  |:|      | |          ||
Db        211 KEDSAEPA-PVEEVGGEVESKPEEKVAVKPESQPSDKPAEESKVEQAGEPV----APRED 265

Qy        169 LQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQ-DME 227
              :    |  :|:  ::|   | :   : |   |:   |||      ||||:|   |  :|
Db        266 EKAPVEPEKQPEAPEEE--KAVEETPKQEESTPDTKAEETVEP--KEETVNQSIEQPKVE 321

Qy        228 EMMSEQENPDSSEPVVEDE----RLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPP 283
                |::   : || ||          |   | |  |:    | | : :   | |  |
Db        322 TPAVEKQTEPTEEPKVEQAGEPVAPREDEQAPTAPVEPEKQPEVPEEEKAV---EETPKP 378

Qy        284 ED--------NPVEDSQV---IVEEVSIFPVE-------------------------EQQ 307
              ||          ||: |::   | :  ]: | :                      :|
Db        379 EDKIKGIGTKEPVDKSELNNQIDKASSVSPTDYSTASYNALGPVLETAKGVYASEPVKQP 438

Qy        308 EVPPETNR-----------KTD------DPEQKAKVKKKKPKLLNKFDKTIKAELDAAEK 350
              ||  |||:           ||:      | : | |          |   :  |||  ||
Db        439 EVNSETNKLKTAIDALNVDKTELNNTIADAKTKVKEHYSDRSWQNLQTEVTKAEKVAANT 498

Qy        351 LRKRGKIEEAVNAF-------------------------KELVRKYPQSPRARYGKAQ 383
              |: :: |||                             :| | ||         | :
Db        499 DAKQSEVNEAVEKLTATIEKLVELSEKPILTLTSTDKKILEREAVAKYTLE---NQNKTK 555

Qy        384 CEDDLAEKRRSNEVLRGAI------------------ETYQ------------------E 407
              :   || ::  ||:   :                   | |:                  |
Db        556 IKSITAELKKGEEVINTVVLTDDKVTTETISAAFKNLEYYKEYTLSTTMIYDRGNGEETE 615

Qy        408 VASLPDVPADLLKLSLK--RRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYL 465
                ::  || |: ||  :|:|  ::          | : |:   |:| |          ||
Db        616 TLENQNIQLDLKKVELKNIKRTDLIKY----ENGKETNESLITTIPDDKSNY------YL 665

Qy        466 LIGDNDN------AKKVYEEVLSVTPNDGFAKVHYGFILK-AQNKIAES-IPYLKEGIES 517
              |  |:        | : |  :: ||       :    : : | ||   |  : |:::
Db        666 KITSNNQKTTLLAVKNIEETTVNGTPVYKVTAIADNLVSRTADNKFEEEYVHYIEK---- 721

Qy        518 GDPGTDDGRFYFHLGDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNV--NGLKAQPW 575
              |  :   |::  : :: : | :  | | ||           |  ||  ||   : :
Db        722 --PKVHEDNVYYNFKELVEAIQN-DPSKEYRLGQ---------SMSARNVVPNG---KSY 766

Qy        576 WTPKETGYTELVKSLERNWKLIRDEG--LAVMDKAKGLF 612
```

```
              | : ||           ||:  ||   |: :    ||
Db        767 ITKEFTG-----------KLLSSEGKQFAITELEHPLF 793
```

RESULT 10
G64242
cytadherence-accessory protein (hmw1) homolog MG386 - Mycoplasma genitalium
C;Species: Mycoplasma genitalium
C;Date: 17-Nov-1995 #sequence_revision 17-Nov-1995 #text_change 07-Dec-1999
C;Accession: G64242
R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.;
M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M
, C.A.; Venter, J.C.
Science 270, 397-403, 1995
A;Title: The minimal gene complement of Mycoplasma genitalium.
A;Reference number: A64200; MUID:96026346; PMID:7569993
A;Accession: G64242
A;Status: preliminary; nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-1616 <TIGR>
A;Cross-references: GB:U39723; GB:L43967; NID:g1046092; PID:g1046097; TIGR:MG386
A;Experimental source: strain G-37
C;Genetics:
A;Genetic code: SGC3

Query Match 4.7%; Score 189.5; DB 2; Length 1616;
Best Local Similarity 22.3%; Pred. No. 0.063;
Matches 142; Conservative 78; Mismatches 205; Indels 213; Gaps 35;

```
Qy         75 FDLVDYEEVLGK----------------LGIYD-------ADGDGDFDVDDAKVL----- 106
              || | :| | |                 : |        : :  |||   |
Db       1091 FDTVKHEAVFDKNQTQTEGLEEPQVSSEAEVVDQTTTDTVGEPEAVFDVQPEKTTEVKFD 1150

Qy        107 -LGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVE 165
               :  :::  ||| |  :  |   || :   ::  ::::|   |  :   :     : ||
Db       1151 DVENQQKVISEPQVEQQPGEAVFEPSAEAKFDSPVESVQDSQPEPVLEEVQTQPEIQPVE 1210

Qy        166 GEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLE-PEVSHEETEHSYHVEETVSQD-CN 223
               :     |  | :|:|   |        |   ||:| || | | |:  |||   | |  |
Db       1211 SQPEATFD--TVQPEQTPQE----AKFDSPVETVEQPEFSSEPTQQ--HVESEASFDEPN 1262

Qy        224 QDMEEMMSEQENP--DS----SEPVVEDERLHHDTDDVTYQV--------YEEQAVYEPL 269
               | :|   : : |  ||    |||  : :  ::| |:  |::         :| |   :|
Db       1263 YDFDEPNYDFDQPSYDSDLQPSEPQYDVDEPNYDFDEPNYEIESKPSEPQFEPQVEQQP- 1321

Qy        270 ENEGIEITEVTAPPE-DNPVE---DSQ--VIVEEVSIFPVEEQQEVPPETNRKTDDPEQK 323
               |  : | :|  : |:|||   |||   ::|||   |  :  |  ||    |  |||
Db       1322 ---GEAVFEPSAEAKFDSPVESVQDSQPEPLLEEVQTQPEIQPVESQPEATFDTVQPEQT 1378

Qy        324 AKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQ 383
               :          |||                                  ||
Db       1379 PQEA--------KFD-----------------------------------SP-------- 1387

Qy        384 CEDDLAEKRRSNEVLRGAIETYQE--VASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLL 441
                                  :|| ||  |:| |:|   ::: : :  |          ::|
Db       1388 ------------------VETIQEPQVSSEPEV---VVQPNFEERKPE---------TVL 1417

Qy        442 TLQRLVQLFP---NDTSLKNDLGVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYGFIL 498
               :  :: |    : ||  :| ||        :|:   ||         ||   | ||
Db       1418 EEPQADEIQPEASEEESLDWELLVG-------NNSYGHYEP-------DG-EWVWAGFFG 1462

Qy        499 KAQ--NKIAESIPYLKEGIESGDPGTDDGRFYFHL-GDAMQ-RVGNKEAYKWYELGHKRG 554
               |  || | :: : :|           | |  | || :  |  ||  : ||    : |
Db       1463 DDQKWNKDA-TVKWARE-----------RDYLPLIGDEVYGRYNNKGEWIWYGFYDESG 1509

Qy        555 HFASV---WQRSLYNVNGLKAQPWWTP----KETGYTELVKSLERNWKLIRDEGLAVMDK 607
               :  |   |:     :|  :|  :|      :| || |      | ||
Db       1510 DWVLVDEQWKNRQPRIN--EAPKFWEKLIGNEEYGYYE---DNEWNW------------- 1551

Qy        608 AKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAP 645
                     :|:|  |   :    ||:  |  |
Db       1552 -------YDGEFDSEGNWLVFQSEETENLNEDITKDIP 1582
```

RESULT 11
T42963
hypothetical protein 48 - ateline herpesvirus 3 (strain 73)
C;Species: ateline herpesvirus 3
A;Variety: strain 73
C;Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jan-2000
C;Accession: T42963
R;Albrecht, J.C.; Fleckenstein, B.
submitted to the EMBL Data Library, August 1998
A;Description: Primary structure of the herpesvirus ateles genome.
A;Reference number: Z22274
A;Accession: T42963
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-792 <ALB>
A;Cross-references: EMBL:AF083424; PIDN:AAC95573.1
A;Experimental source: strain 73

Query Match 4.7%; Score 188.5; DB 2; Length 792;
Best Local Similarity 21.8%; Pred. No. 0.028;
Matches 90; Conservative 58; Mismatches 147; Indels 117; Gaps 14;

```
Qy          94 GDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQS 153
               |||: | :| :                 || |   | ||:   | | :  |||  |:
Db         464 GDGEEDEEDEE-----------------EEDEEEEEEEEEEEDEEEEEEEEDEEDEE--- 503

Qy         154 LLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYH 213
                 |  : || : |:    |  :||:|     : ::  |  | |   || |
Db         504 --------EEEDEEDEEDEEDEEDEEDEEDEEDEEEEEDEEEEDEDEEEEEEEEE 555

Qy         214 VEETVSQDCNQDMEEMMSEQENPDSSEPV-VEDERLHHDTDDVTYQVYEEQAVYEPLENE 272
                ||   ::  :| ||   |:|  |  |    |||    | :|      ||:      | |
Db         556 EEEEEDEEDEEDEEEKEDEEEKEDEEEKEDEEDEEEKEDDED------EEE------EEE 603

Qy         273 GIEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPK 332
               | |  :    |:   || :  :|     ||: |   : : : :| | : :  ::  :
Db         604 GEEKEDDEDDEEEEDEEDDEEEEDE------EEEDEDEEDEDEEDEDEEDEDEEDEEGDR 657

Qy         333 LLNKFDKTIKAELDAAE----------KLRKRGKIEEAVNAFKELVRKYPQSPRARYGKA 382
                   |:   |: |  :           |  ::|      :::|
Db         658 -----DRDADADRDGGDGDGDGVGYDYKDEEKG-----TDSYKN---------------- 691

Qy         383 QCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLT 442
                 | :   | : :| | |  : |: |   |       |   :  ||   :   :| |:|:
Db         692 --ETEYNNKEQKDENLHGESKYYEPVFFKP-------KQHTENLSD---YYNWLRNSMLS 739

Qy         443 LQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHY 494
                       || |  :          ::: : |   |  | | | |: |
Db         740 -----------TSDKKSI-----------SSEDILENNSSYLPIDHFEKILY 769
```

RESULT 12
A45990
junctional sarcoplasmic reticulum glycoprotein triadin - rabbit
C;Species: Oryctolagus cuniculus (domestic rabbit)
C;Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 11-Jan-2000
C;Accession: A45990; S45704
R;Knudson, C.M.; Stang, K.K.; Moomaw, C.R.; Slaughter, C.A.; Campbell, K.P.
J. Biol. Chem. 268, 12646-12654, 1993
A;Title: Primary structure and topological analysis of a skeletal muscle-specific juncti
A;Reference number: A45990; MUID:93286104; PMID:7685347
A;Accession: A45990
A;Status: preliminary
A;Molecule type: mRNA; protein
A;Residues: 1-706 <KNU>
A;Cross-references: GB:L10065; NID:g347850; PIDN:AAA31488.1; PID:g347851
A;Experimental source: triads, muscle
A;Note: sequence extracted from NCBI backbone (NCBIN:133603, NCBIP:133604)
R;Peng, M.; Fan, H.; Kirley, T.L.; Caswell, A.H.; Schwartz, A.
FEBS Lett. 348, 17-20, 1994
A;Title: Structural diversity of triadin in skeletal muscle and evidence of its existenc
A;Reference number: S45704; MUID:94298946; PMID:8026576
A;Accession: S45704
A;Molecule type: mRNA
A;Residues: 401-620 <PEN>
C;Superfamily: histone H1
C;Keywords: glycoprotein

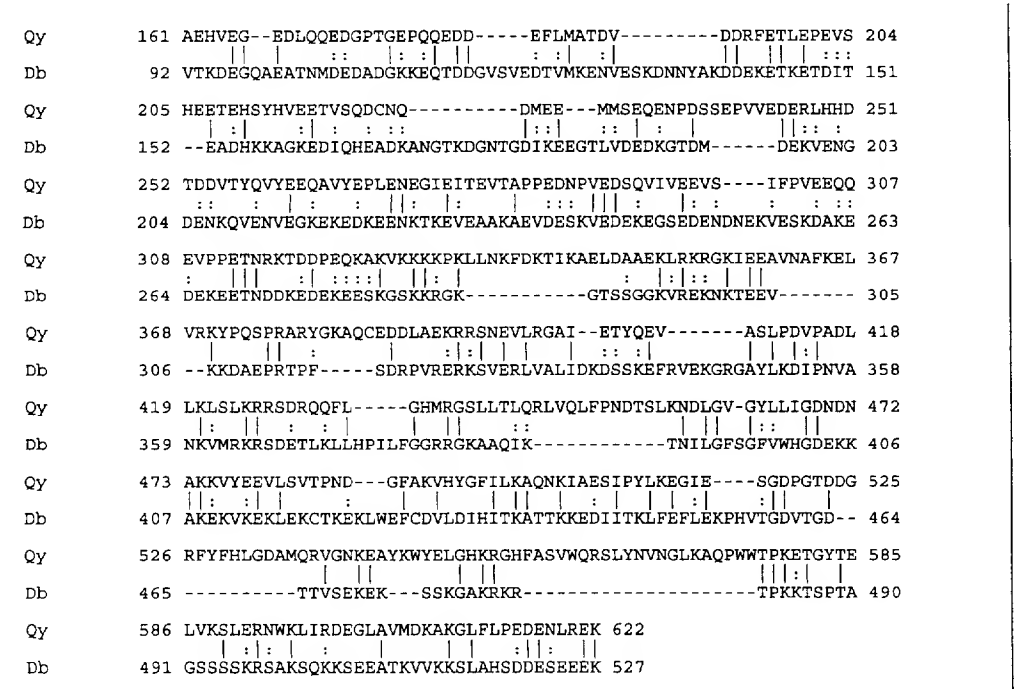
Query Match 4.6%; Score 187; DB 2; Length 706;
Best Local Similarity 21.0%; Pred. No. 0.029;
Matches 105; Conservative 77; Mismatches 165; Indels 152; Gaps 22;

```
Qy           6 NAKSSGNSSSSGSGSGSTSAGS--------------SSPGARRETKHGGHKNGRKGGLSG 51
                 || ::     | :||   | |                |||  |
Db           9 NASTTTTVIDSKNGSVPKSPGKVLKRTVTEDLVTTFSSPAA------------------- 49

Qy          52 TSFFTWFMVIALLGVWTSVAVVWFDLVDYEEV----LGKLG------IYDA--------- 92
                  | :||||:  |::|||| ||||||:      : |:|      ::||
Db          50 -----WLLVIALIITWSAVAVVMFDLVDYKNFSASSIAKMGSDPLKLVHDAVEETTDWIY 104

Qy          93 ------------DGDGDFDVDDAKVLLG-LKERSTSEPAVPPEEAEPHTEPEEQVPV--- 136
                         ||| : |  |     | ::|       :  |: |   :|| ::|
Db         105 GFFSLLSDIISSDGDEEDDEGDEDTAKGEIEEPPLKRKDIHKEKIEKQEKPERKIPTKVV 164

Qy         137 ----EAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDV 192
                    | | : :::: | :      :  | | :| ::  :    | | ::   :
Db         165 HKEKEKEKEKVKEKEKPE-----KKATHKEKLEKKEKPETKTVTKEEKKARTK------- 212

Qy         193 DDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDT 252
                    | :|  |  |:  :  | : :   |: |          |:   :|
Db         213 ----EKIEEKTKKE-------VKGVKQEKVKQTVAKAKEVQKTPKPK------EKESKET 255

Qy         253 DDVTYQVYEEQAVY-----EPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQ 307
                  |: |  ::|  :     :  :  ::  :  | |  :| | :  |          |::
Db         256 AAVSKQEQKDQYAFCRYMIDIFVHGDLKPGQSPAIPPPSPTEQAS---RPTPALPTPEEK 312

Qy         308 E---------VPPETNRKTDDPEQKAK--------VKKKKP-KLLNKFDKTIKAELDAAE 349
                |         |  || :| :  : | |         :|||:| |  :    |:|    ||
Db         313 EGEKKKAEKKVTTETKKKAEKEDAKKKSEKETDIDMKKKEPGKSPDTKPGTVKVTTQAAT 372

Qy         350 KLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVA 409
                | |  | |:::  | |:   : |:      || |       ||:  :|    | | :| |
Db         373 K--KDEKKEDSKKA-KKPAEEQPK------GKKQ------EKKEKHE--EPAKSTKKEHA 415

Qy         410 SLPDVPADLLKLSLKRRSD 428
               :  :  |   |  ::|: :
Db         416 APSEKQA---KAKIERKEE 431
```

RESULT 13
T08929
hypothetical protein T15N24.80 - Arabidopsis thaliana
C;Species: Arabidopsis thaliana (mouse-ear cress)
C;Date: 11-Jun-1999 #sequence_revision 11-Jun-1999 #text_change 22-Oct-1999
C;Accession: T08929
R;Bevan, M.; Zimmermann, W.; Grueneisen, A.; Wambutt, R.; Bancroft, I.; Mewes, H.W.; May
submitted to the Protein Sequence Database, May 1999
A;Reference number: Z16518
A;Accession: T08929
A;Molecule type: DNA
A;Residues: 1-763 <BEV>
A;Cross-references: EMBL:AL078465; GSPDB:GN00062; ATSP:T15N24.80
A;Experimental source: cultivar Columbia; BAC clone T15N24
C;Genetics:
A;Gene: ATSP:T15N24.80
A;Map position: 4
A;Introns: 345/3; 357/1; 384/3; 404/3; 441/3; 472/3; 500/3; 667/1; 697/3; 708/1

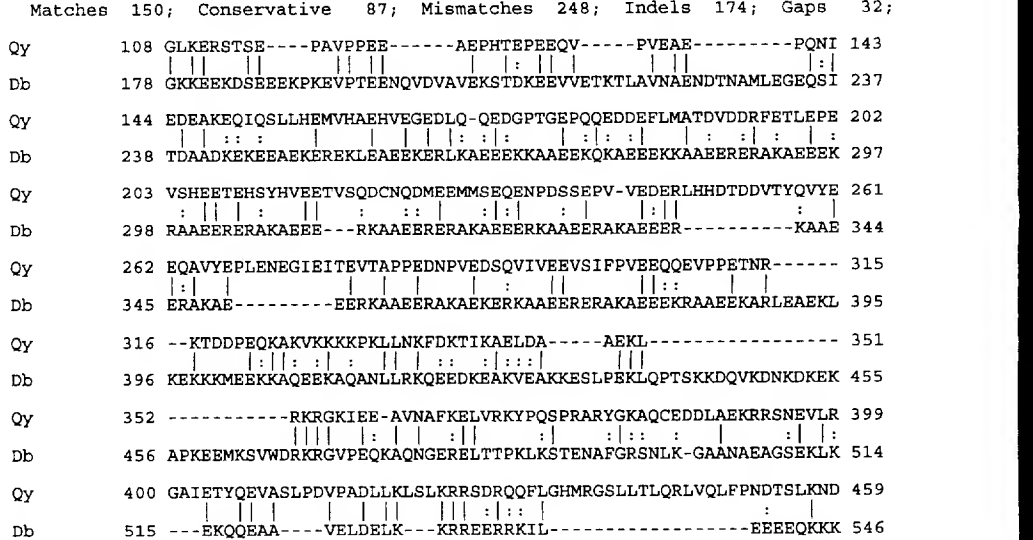
Query Match 4.5%; Score 183; DB 2; Length 763;
Best Local Similarity 21.8%; Pred. No. 0.052;
Matches 126; Conservative 84; Mismatches 217; Indels 150; Gaps 23;

```
Qy         112 RSTSEPAVPPEEAEPHT----------EPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVH 160
               : | | |   : |||              |  |    |:| :  || |: |    : | |
Db          35 KETQELAKDEDMAEPDNMEIDAQIKKDDEKAETEDKESEVKKNEDNAETQ---KMEEKVE 91
```



RESULT 14
A33430
h-caldesmon - chicken
N;Alternate names: caldesmon, smooth muscle; calmodulin- and actin-binding protein
C;Species: Gallus gallus (chicken)
C;Date: 27-Feb-1990 #sequence_revision 27-Jun-1994 #text_change 22-Jun-1999
C;Accession: A33430; A32642; A32445; A41064; A60461; PC2003; PX0022
R;Hayashi, K.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
Biochem. Biophys. Res. Commun. 164, 503-511, 1989
A;Title: Primary structure and functional expression of h-caldesmon complementary DNA.
A;Reference number: A33430; MUID:90026426; PMID:2803315
A;Accession: A33430
A;Molecule type: mRNA
A;Residues: 1-771 <HAY>
A;Cross-references: GB:M28417; NID:g211895; PIDN:AAA48810.1; PID:g211896
A;Experimental source: gizzard
A;Note: part of this sequence was confirmed by protein sequencing
R;Bryan, J.; Imai, M.; Lee, R.; Moore, P.; Cook, R.G.; Lin, W.G.
J. Biol. Chem. 264, 13873-13879, 1989
A;Title: Cloning and expression of a smooth muscle caldesmon.
A;Reference number: A32642; MUID:89340480; PMID:2760048
A;Accession: A32642
A;Molecule type: mRNA
A;Residues: 1-318,334-771 <BRY>
A;Cross-references: GB:J04968; NID:g212656; PIDN:AAA49067.1; PID:g212657
A;Note: the authors translated the codon GAA for residue 743 as Leu
A;Note: this alternative splice form is a high molecular weight caldesmon (h-caldesmon)
R;Hayashi, K.; Yamada, S.; Kanda, K.; Kimizuka, F.; Kato, I.; Sobue, K.
Biochem. Biophys. Res. Commun. 161, 38-45, 1989
A;Title: 35kDa fragment of h-caldesmon conserves two consensus sequences of the tropomyo
A;Reference number: A32445; MUID:89273666; PMID:2730665
A;Accession: A32445
A;Molecule type: mRNA
A;Residues: 466-771 <HA2>
A;Cross-references: GB:M26684; NID:g211897; PIDN:AAA48811.1; PID:g211898
R;Mak, A.S.; Carpenter, M.; Smillie, L.B.; Wang, J.H.
J. Biol. Chem. 266, 19971-19975, 1991
A;Title: Phosphorylation of caldesmon by p34(cdc2) kinase. Identification of phosphorylat
A;Reference number: A41064; MUID:92041815; PMID:1939059
A;Accession: A41064
A;Molecule type: protein
A;Residues: 597-600;678-696;711-721 <MAK>
R;Wang, C.L.A.; Wang, L.W.C.; Lu, R.C.
Biochem. Biophys. Res. Commun. 162, 746-752, 1989
A;Title: Caldesmon has two calmodulin-binding domains.
A;Reference number: A60461; MUID:89334885; PMID:2757638
A;Accession: A60461
A;Molecule type: protein
A;Residues: 2-17,'X',19-38;466-485 <WAN>
R;Haruna, M.; Hayashi, K.; Yano, H.; Takeuchi, O.; Sobue, K.
Biochem. Biophys. Res. Commun. 197, 145-153, 1993
A;Title: Common structural and expressional properties of vertebrate caldesmon genes.
A;Reference number: PC2003; MUID:94071934; PMID:8250919
A;Accession: PC2003
A;Molecule type: DNA
A;Residues: 74-419 <HAR>
R;Takagi, T.; Yazawa, M.; Ueno, T.; Suzuki, S.; Yagi, K.
J. Biochem. 106, 778-783, 1989
A;Title: Amino acid sequence studies on cyanogen bromide peptides of chicken caldesmon wh
A;Reference number: PX0022; MUID:90130380; PMID:2613684
A;Accession: PX0022
A;Molecule type: protein
A;Residues: 462-477,'D',479-563;674-762,'A',763-771 <TAK>
C;Comment: This protein plays a vital role in the regulation of smooth muscle and nonmusc
C;Comment: The binding of caldesmon to F-actin is modulated by calcium and calmodulin.
C;Comment: Two calmodulin molecules can bind to nonoverlapping domains of each caldesmon.
C;Superfamily: caldesmon
C;Keywords: actin binding; alternative splicing; calmodulin binding; cytoskeleton; muscle
F;1-771/Product: h-caldesmon #status predicted <HMAT>
F;1-318,334-771/Product: h-caldesmon, alternative splice form #status predicted <LMAT>
F;266-390/Region: 13-residue repeats
F;511-582/Region: tropomyosin binding
F;622-636/Region: tropomyosin binding
F;597,682,717/Binding site: phosphate (Ser) (covalent) (by cdc2 kinase) #status experimer
F;688,711/Binding site: phosphate (Thr) (covalent) (by cdc2 kinase) #status experimental

Query Match 4.5%; Score 181; DB 1; Length 771;
Best Local Similarity 22.8%; Pred. No. 0.066;
Matches 150; Conservative 87; Mismatches 248; Indels 174; Gaps 32;



```
Qy        460 LGVGYLLIGDNDNAKKVYEEV----------LSVTPNDGFA---KVHYGFILKAQN-KIA 505
Db        547 QEEAERKIREEEEKKRMKEEIERRRAEAAEKRQKVPEDGVSEEKKPFKCFSPKGSSLKIE 606

Qy        506 ESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGNK-EAYKWYELGHKRG---------- 554
Db        607 ERAEFLNKSAQK------SGMKPAHTTAVVSKIDSRLEQYTSAVVGNKAAKPAKPAASDL 660

Qy        555 --------HFASVWQR-SLYNVNGLKAQPWWTPKET-GYTELVKSLERNWKLIRDEGLAV 604
Db        661 PVPAEGVRNIKSMWEKGNVFSSPGGTGTP---NKETAGLKVGVSSRINEWLTKTPEG--- 714

Qy        605 MDKAKGLFLPEDENLREKGDWS-QFTLWQQGRRNENACKGAPKTCTLLEKFPETTGCRR 662
Db        715 -NKSPA---PKPSDLR-PGDVSGKRNLWEKQSVEKPAASSSKVTAT--GKKSETNGLRQ 766


RESULT 15
S47436
flagellar antigen - Trypanosoma brucei (fragment)
C;Species: Trypanosoma brucei
C;Date: 13-Jan-1995 #sequence_revision 30-Jan-1998 #text_change 03-Nov-2000
C;Accession: S47436
R;Imboden, M.; Mueller, N.; Hemphill, A.; Mattioli, X.Y.Z.; Seebeck, T.
submitted to the EMBL Data Library, August 1994
A;Description: Repetitive proteins from the flagellar cytoskeleton of African Trypanosom
A;Reference number: S47436
A;Accession: S47436
A;Molecule type: mRNA
A;Residues: 1-411 <IMB>
A;Cross-references: EMBL:Z36281; NID:g530358; PID:g530359
A;Experimental source: strain stock TREU 1285
C;Superfamily: cytadherence-accessory protein hmw1

  Query Match             4.5%;  Score 179.5;  DB 2;  Length 411;
  Best Local Similarity   25.1%;  Pred. No. 0.036;
  Matches   83;  Conservative   53;  Mismatches  128;  Indels   67;  Gaps   16;

Qy        116 EPAVPPEEAEPHTEPEEQVPVEA-----EPQNIEDEAKEQIQSLLHEMVHAE-HVEGEDL 169
Db         11 EPQQVPAEAQPEAQPEGDIAVEALEELEEPQQVPAEAQP-------EAVAPEGDIAVEAL 63

Qy        170 QQEDGPTGEPQQEDDEFLMATDVD-DRFETLE--PEVSHEETEHSYHVEETVSQDCNQDM 226
Db         64 EELEEPQQAPAEAQPEALPEGDIAVEALEELEEPQQVPAEAQPEAVAPEGDIAVEALEEL 123

Qy        227 EEMMSEQENPDSSEP-VVEDERLHHDTDDVTYQVYEE---------QAVYEPLENEG--- 273
Db        124 EE---PQQVPAEAQPEAVAPE------GDIAVEALEELEEPQQVPAEAQPEAVAPEGDIA 174

Qy        274 IEITEVTAPPEDNPVE-------DSQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKV 326
Db        175 VEALEELEEPQQVPAEAQPEAQPEGDIAVE--ALEELEEPQQVPAEAQPEAVAPEGDIAV 232

Qy        327 KKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCED 386
Db        233 -----EALEELEEPQQAPAEAQPEAQPEGDI--AVEALEEL-EEPQQAPAEAQPEAQPEG 284

Qy        387 DLAEKRRSNEVLRGAIETYQEVASLPDVPAD 417
Db        285 DI------------AVEALEELEEPQQAPAE 303


Search completed: May  5, 2004, 11:04:29
Job time : 23 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 5, 2004, 10:56:23 ; Search time 18 Seconds
(without alignments)
2192.732 Million cell updates/sec

Title: US-09-903-216-2
Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSSGSGS.....IVDVWHPELTPQQRRSLPAI 758

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	3966.5	98.6	757	1	ASPH_HUMAN	Q12797 homo sapien
2	3172	78.9	754	1	ASPH_BOVIN	Q28056 bos taurus
3	203	5.0	700	1	TRDN_CANFA	P82179 canis famil
4	194	4.8	4835	1	MDN1_GIALA	Q8t5t1 giardia lam
5	193	4.8	493	1	ECX1_METMA	Q8ptt8 methanosarc
6	189.5	4.7	1616	1	P200_MYCGE	Q49429 mycoplasma
7	187	4.6	705	1	TRDN_RABIT	Q28820 oryctolagus
8	181	4.5	771	1	CALD_CHICK	P12957 gallus gall
9	178.5	4.4	728	1	TRDN_HUMAN	Q13061 homo sapien
10	176.5	4.4	845	1	NFM_RAT	P12839 rattus norv
11	175	4.4	2004	1	MYS3_HUMAN	Q92794 homo sapien
12	172.5	4.3	1240	1	YNJ1_YEAST	P53935 saccharomyc
13	172	4.3	465	1	YHU6_YEAST	P38845 saccharomyc
14	172	4.3	630	1	YCF2_OENVI	P31569 oenothera v
15	169.5	4.2	592	1	LAM2_MOUSE	P21619 mus musculu
16	169.5	4.2	1233	1	YFI6_YEAST	P43597 saccharomyc
17	165	4.1	650	1	NUCL_XENLA	P20397 xenopus lae
18	164	4.1	1957	1	SPOF_SCHPO	Q10411 schizosacch
19	163	4.1	411	1	FKB3_YEAST	P38911 saccharomyc
20	162.5	4.0	793	1	CALD_HUMAN	Q05682 homo sapien
21	161.5	4.0	848	1	NFM_MOUSE	P08553 mus musculu
22	161.5	4.0	1220	1	IF2P_HUMAN	O60841 homo sapien
23	161.5	4.0	1395	1	SP41_YEAST	P38904 saccharomyc
24	161	4.0	2663	1	CENE_HUMAN	Q02224 homo sapien
25	161	4.0	4910	1	MDN1_YEAST	Q12019 saccharomyc
26	160.5	4.0	721	1	YCF2_OENPI	P31568 oenothera p
27	159	4.0	3924	1	ANK2_HUMAN	Q01484 homo sapien
28	158.5	3.9	837	1	RA50_METTH	O26640 methanobact
29	158.5	3.9	1102	1	YG49_SCHPO	O60184 schizosacch
30	158	3.9	795	1	CDL1_HUMAN	P21127 homo sapien
31	157.5	3.9	780	1	CDL2_HUMAN	Q9uq88 homo sapien
32	157.5	3.9	1357	1	KTN1_HUMAN	Q86up2 homo sapien
33	156.5	3.9	506	1	NPL3_HUMAN	Q99457 homo sapien
34	156.5	3.9	1658	1	YM67_YEAST	Q03661 saccharomyc
35	154.5	3.8	1828	1	MAP2_MOUSE	P20357 mus musculu
36	154	3.8	1549	1	TRHY_SHEEP	P22793 ovis aries
37	153.5	3.8	1898	1	TRHY_HUMAN	Q07283 homo sapien
38	152	3.8	434	1	YKI2_YEAST	P36080 saccharomyc
39	152	3.8	671	1	CHS5_YEAST	Q12114 saccharomyc
40	152	3.8	725	1	HS9B_BRARE	O57521 brachydanio
41	152	3.8	728	1	HS9A_CHICK	P11501 gallus gall
42	152	3.8	1001	1	RPGR_MOUSE	Q9r0x5 mus musculu
43	152	3.8	1233	1	SM1A_BOVIN	O97593 bos taurus
44	152	3.8	1233	1	SM1A_HUMAN	Q14683 homo sapien
45	152	3.8	1282	1	BMS1_HUMAN	Q14692 homo sapien

ALIGNMENTS

RESULT 1
ASPH_HUMAN
ID ASPH_HUMAN STANDARD; PRT; 757 AA.
AC Q12797;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Aspartyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-
DE hydroxylase) (ASP beta-hydroxylase) (Peptide-aspartate beta-
DE dioxygenase).
GN ASPH.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95121937; PubMed=7821814;
RA Korioth F., Gieffers C., Frey J.;
RT "Cloning and characterization of the human gene encoding aspartyl
RT beta-hydroxylase.";
RL Gene 150:395-399(1994).
CC -!- FUNCTION: Specifically hydroxylates an Asp or Asn residue in
CC certain epidermal growth factor-like (EGF) domains of a number of
CC proteins.
CC -!- CATALYTIC ACTIVITY: Peptide L-aspartate + 2-oxoglutarate + O(2) =
CC peptide 3-hydroxy-L-aspartate + succinate + CO(2).
CC -!- COFACTOR: Iron.
CC -!- SUBUNIT: Monomer (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC reticulum.
CC -!- TISSUE SPECIFICITY: Detected in all tissues tested.
CC -!- PTM: MIGHT BE PROCESSED TO THE 56 kDa (AA 274-757) OR 52 kDa (AA
CC 315-757) FORMS IN THE LUMEN OF THE ENDOPLASMIC RETICULUM (BY
CC SIMILARITY).
CC ----------

CC ----------
DR EMBL; U03109; AAA82108.1; -.
DR PIR; I38423; I38423.
DR Genew; HGNC:757; ASPH.
DR MIM; 600582; -.
DR GO; GO:0005789; C:endoplasmic reticulum membrane; TAS.
DR GO; GO:0005509; F:calcium ion binding; TAS.
DR GO; GO:0005489; F:electron transporter activity; TAS.
DR GO; GO:0004597; F:peptide-aspartate beta-dioxygenase activity; TAS.
DR GO; GO:0008307; F:structural constituent of muscle; TAS.
DR GO; GO:0006936; P:muscle contraction; TAS.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.

```
DR   InterPro; IPR008940; Prenyl_trans.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
DR   Pfam; PF05118; Asp_Arg_Hydrox; 1.
KW   Oxidoreductase; Dioxygenase; Iron; Transmembrane; Signal-anchor;
KW   Endoplasmic reticulum.
FT   DOMAIN        1     54       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     55     75       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       76    757       LUMENAL (POTENTIAL).
FT   DOMAIN       13     20       POLY-SER.
FT   DOMAIN      323    332       POLY-LYS.
FT   CARBOHYD    452    452       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    705    705       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   757 AA;  85498 MW;  1A79313A4934C430 CRC64;

  Query Match             98.6%;  Score 3966.5;  DB 1;  Length 757;
  Best Local Similarity   99.2%;  Pred. No. 3.4e-193;
  Matches  752;  Conservative    2;  Mismatches    3;  Indels    1;  Gaps    1;

Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              |||||||||||||||||||||| ||||||||   |||||||:|||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLINVNGLKAQP-CGPKETGYTQLVKSLERNWKLIRDE 599

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        600 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 659

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        660 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETKTWEEGKVLIFD 719

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||||
Db        720 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 757
```

```
RESULT 2
ASPH_BOVIN
ID   ASPH_BOVIN     STANDARD;      PRT;   754 AA.
AC   Q28056;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Aspartyl/asparaginyl beta-hydroxylase (EC 1.14.11.16) (Aspartate beta-
DE   hydroxylase) (ASP beta-hydroxylase) (Peptide-aspartate beta-
DE   dioxygenase).
GN   ASPH.
OS   Bos taurus (Bovine).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC   Bovidae; Bovinae; Bos.
OX   NCBI_TaxID=9913;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Brain, and Liver;
RX   MEDLINE=92332546; PubMed=1378441;
RA   Jia S., Vandusen W.J., Diehl R.E., Kohl N.E., Dixon R.A.F.,
RA   Elliston K.O., Stern A.M., Friedman P.A.;
RT   "cDNA cloning and expression of bovine aspartyl (asparaginyl) beta-
RT   hydroxylase.";
RL   J. Biol. Chem. 267:14322-14327(1992).
RN   [2]
RP   SEQUENCE OF 289-385 AND 615-641.
RC   TISSUE=Liver;
RX   MEDLINE=91310689; PubMed=1856229;
RA   Wang Q., Vandusen W.J., Petroski C.J., Garsky V.M., Stern A.M.,
RA   Friedman P.A.;
RT   "Bovine liver aspartyl beta-hydroxylase. Purification and
RT   characterization.";
RL   J. Biol. Chem. 266:14004-14010(1991).
CC   -!- FUNCTION: Specifically hydroxylates an Asp or Asn residue in
CC       certain epidermal growth factor-like (EGF) domains of a number of
CC       proteins.
CC   -!- CATALYTIC ACTIVITY: Peptide L-aspartate + 2-oxoglutarate + O(2) =
CC       peptide 3-hydroxy-L-aspartate + succinate + CO(2).
CC   -!- COFACTOR: Iron.
CC   -!- SUBUNIT: Monomer.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Endoplasmic
CC       reticulum.
CC   -!- PTM: Might be processed to the 56 kDa (AA 289-754) or 52 kDa (AA
CC       311-754) forms in the lumen of the endoplasmic reticulum.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M91213; AAA03563.1; -.
DR   PIR; A42969; BABOH.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   InterPro; IPR007803; Asp_Arg_Hydrox.
DR   InterPro; IPR008941; TPR-like.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
DR   Pfam; PF05118; Asp_Arg_Hydrox; 1.
KW   Oxidoreductase; Dioxygenase; Iron; Transmembrane; Signal-anchor;
KW   Endoplasmic reticulum.
FT   DOMAIN        1     57       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     58     78       SIGNAL-ANCHOR (TYPE-II MEMBRANE PROTEIN)
FT                                (POTENTIAL).
FT   DOMAIN       79    754       LUMENAL (POTENTIAL).
FT   DOMAIN        9     12       POLY-GLY.
FT   DOMAIN       14     21       POLY-SER.
FT   DOMAIN      318    328       POLY-LYS.
FT   CARBOHYD     96     96       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    466    466       N-LINKED (GLCNAC. . .) (POTENTIAL).
```

```
FT   CARBOHYD    702    702       N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ   SEQUENCE   754 AA;  84998 MW;  369593A1F0B558C8 CRC64;

  Query Match             78.9%;  Score 3172;  DB 1;  Length 754;
  Best Local Similarity   78.7%;  Pred. No. 4.2e-153;
  Matches  612;  Conservative   45;  Mismatches   77;  Indels   44;  Gaps    6;

Qy          1 MAQRKNAKSSG---NSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTW 57
              || |||||  |   :|||||| :| || ||||||||||||| || ||||||||||:|||||
Db          1 MAPRKNAKGGGGNSSSSSSGSPTGCTSGGSSSPGARRETKQGGLKNGRKGGLSGSSFFTW 60

Qy         58 FMVIALLGVWTSVAVVWFDLVDYEEVL--------------GKLGIYDADGDGDFDVDD 102
              ||||||||||||||||||||||||||||                ||||||||||||||||||
Db         61 FMVIALLGVWTSVAVVWFDLVDYEEVLAKAKDFRYNLSEVLQGKLGIYDADGDGDFDVDD 120

Qy        103 AKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAE 162
              |||||||||:   :| ||||||: :   |:||       ||||||  ||:|| | | |::|
Db        121 AKVLLGLKEKPAPKPTVPPEEADMYPWLEDQVLESPGRQNIEDEVYEQVQS-LDETVYSE 179

Qy        163 HVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDC 222
              ||:| ||    |  | ||   : :| |||:| :     ||||| |||:||| |
Db        180 --PGENLPQEPEGPAEELQPDDHVFVGSDADDRYEPMGTGAVHEETEDSYHIEETASPAY 237

Qy        223 NQDMEEMMSEQENPDSSEPVVED--ERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVT 280
              :||||:|| |||||||||||| |  || : :||||||: |:||
Db        238 SQDMEDMMYEQENPDSSEPVVVDDAERTYQETDDVTYRDYDEQ---------------- 280

Qy        281 APPEDNPVEDSQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKT 340
                  |: |::|   |:||  : | |||||||||:| |:| :| |||||||||||||||
Db        281 ----DHAVDNSNTILEEPHMPPAEEQQEVPPETNKKADEPGKKGKVKKKKPKLLNKFDKT 336

Qy        341 IKAELDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRG 400
              ||||||||||||||||||||||:|||||||||| |||||||||||||||||||||:||
Db        337 IKAELDAAEKLRKRGKIEEAVNAFEELVRKYPQSPGARYGKAQCEDDLAEKRRSNEILRR 396

Qy        401 AIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDL 460
              ||||||| ||||| | ||:|||||||||||||||||||||:||||||:||:|||||
Db        397 AIETYQEAASLPDAPTDLVKLSLKRRSDRQQFLGHMRGSLLTLQKLVQLFPDDTALKNDL 456

Qy        461 GVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDP 520
              ||||||||||:|||||||||||||||||||||||||||||||||||||||||||||||||
Db        457 GVGYLLIGDNDSAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDP 516

Qy        521 GTDDGRFYFHLGDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKE 580
              ||||||||||||||||||||||:||||||:||||||||||||| ||||||||||||
Db        517 GTDDGRFYFHLGDAMQRVGNKEAYRWYELGHQRGHFASVWQRSLYNVQGLKAQPWWTPKE 576

Qy        581 TGYTELVKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENA 640
              |||||||||||||||||||||| ||:  |||||||||||||||||||||||:||||
Db        577 TGYTELVKSLERNWKLIRDEGLAAMDRTHGLFLPEDENLREKGDWSQFTLWQQGRKNENA 636

Qy        641 CKGAPKTCTLLEKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGC 700
              |||||||||:||:|||||||||||||||||||||||||||||||||||||||||||||||
Db        637 CKGAPKTCSLLDKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGC 696

Qy        701 KIRCANETRTWEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||:|||||||||||||||| ||||||||
Db        697 KIRCANETRTWEEGKVLIFDDSFEHEVWQDAASFRLIFIVDVWHPELTPHQRRSLPAI 754


RESULT 3
TRDN_CANFA
ID   TRDN_CANFA     STANDARD;      PRT;   700 AA.
AC   P82179;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Triadin.
GN   TRDN.
OS   Canis familiaris (Dog).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX   NCBI_TaxID=9615;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   TISSUE=Heart, and Skeletal muscle;
RX   MEDLINE=99428545; PubMed=10497235;
RA   Kobayashi Y.M., Jones L.R.;
RT   "Identification of triadin 1 as the predominant triadin isoform
RT   expressed in mammalian myocardium.";
RL   J. Biol. Chem. 274:28660-28668(1999).
CC   -!- FUNCTION: May be involved in anchoring calsequestrin to the
CC       junctional sarcoplasmic reticulum and allowing its functional
CC       coupling with the ryanodine receptor (By similarity).
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC       reticulum.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=Skeletal;
CC         IsoId=P82179-1; Sequence=Displayed;
CC       Name=Cardiac 1;
CC         IsoId=P82179-2; Sequence=VSP_004001, VSP_004002;
CC       Name=Cardiac 3;
CC         IsoId=P82179-3; Sequence=VSP_004003, VSP_004004;
CC   -!- TISSUE SPECIFICITY: Skeletal and cardiac muscle.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF165916; AAF00222.1; -.
DR   EMBL; AF165915; AAF00221.1; -.
DR   EMBL; AF165917; AAF00223.1; -.
KW   Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW   Alternative splicing.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1     46       CYTOPLASMIC (POTENTIAL).
FT   TRANSMEM     47     67       POTENTIAL.
FT   DOMAIN       68    700       LUMENAL (POTENTIAL).
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .).
FT   CARBOHYD    616    616       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   VARSPLIC    257    277       DQYAFCRYMIDMFVHGDLRPG -> GKHSEEVAGGSKRTLG
FT                                KKQIQ (in isoform Cardiac 1).
FT                                /FTId=VSP_004001.
FT   VARSPLIC    278    700       Missing (in isoform Cardiac 1).
FT                                /FTId=VSP_004002.
FT   VARSPLIC    466    466       E -> EPIKGKEVKVPGSLKEKE (in isoform
FT                                Cardiac 3).
FT                                /FTId=VSP_004003.
FT   VARSPLIC    524    579       EEKVVKQVKATEKAAIEKTVKPKPAKKAEHQEKESPTIKTD
FT                                KPKPTSKETPEVTES -> GILQVVPVVLNCLFLVQFQQDE
FT                                ELNVESKVFRMIHVLSHPTSRTSPILVISTTCRT (in
FT                                isoform Cardiac 3).
FT                                /FTId=VSP_004004.
SQ   SEQUENCE   700 AA;  78152 MW;  F033E3AA1BEE0C56 CRC64;

  Query Match             5.0%;  Score 203;  DB 1;  Length 700;
  Best Local Similarity   19.3%;  Pred. No. 0.0027;
  Matches  141;  Conservative   95;  Mismatches  236;  Indels  258;  Gaps   28;

Qy          6 NAKSSGNSSSSGSGSGSTSAGS--------------SSPGARRETKHGGHKNGRKGGLSG 51
              || ::      | :||    | |                 ||| |
Db          8 NASTTTTVIDSKNGSVPKSPGKVLKRTVTEDIVTTFSSPAA------------------- 48

Qy         52 TSFFTWFMVIALLGVWTSVAVVWFDLVDYEEV----LGKLG------IYDA--------- 92
                   | :||||:  |::|||| ||||||:       | |:|       ::||
Db         49 -----WLLVIALIITWSAVAVVMFDLVDYKNFSASSLSKIGSDPLKLVHDAVEETTDWVY 103

Qy         93 ----------------DGDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVP 135
                            | ||| | |  ::    :|    :  :  |:||   :|| ::
Db        104 GFFSLLSDIISSDGDEDDDDGDEDTDKGEI----EEPPLKQKEIHKEKAEKEEKPERKIL 159
```

```
Qy        136 VEA---EPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDV 192
Db        160 AKVAHREKEKVKEKEKSE-----KKATHKEKIEKKEKPETKTMAKEERKAKTEEKIKKEV 214

Qy        193 -DDRFETLEPEVS------------------------HEETEH----SYHVEETVSQDC 222
Db        215 KGGKQEKVKPTAAKVKEVQKTPPKAKEKEGKETAAVAKHEQKDQYAFCRYMIDMFVHGDL 274

Qy        223 NQDMEEMM--------------------SEQENPDSSEPVVEDERLHHDTDDVTYQVYE 261
Db        275 RPGQSPALPPPLPTVQASRPTPASPTLEGKEEEEKKKAEKKVTSETKKKEKEDVKKKSDK 334

Qy        262 EQAV--------YEPLENEG-IEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQE---- 308
Db        335 DTAIDVEKKEPGKAPETKQGTIKVVAQAAAKKDEKKEDSKKTKTPVEEHPKGKKQEKKEK 394

Qy        309 -VPPETNRKTDD--PEQK--------------AKVKKKKP-KLLNKFDKTIKAELDAAE 349
Db        395 YVEPAKSSKKEHSAPSEKQVKAKTERAKEETSAASTKKAVPGKKEEKTTKTVEQEI---- 450

Qy        350 KLRKRGKI--------------EEAVNAFKELVRKYPQSPRARYGK--------AQCEDD 387
Db        451 RKEKSGKTSTASKDKEPEIKKDEKMPKADKEVKPKPPQSQVKKEEKSESQVKKEAKPEQD 510

Qy        388 LAEKRRS--------------NEVLRGAIE-----------TYQEVASLPDVPADLLKLS 422
Db        511 IAKPEKTVSHGKPEEKVVKQVKATEKAAIEKTVKPKPAKKAEHQEKES-PTIKTDKPKPT 569

Qy        423 LK-----------------RRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYL 465
Db        570 SKETPEVTESGKKKIEKSEKESKEKAEMKHLKEEKVSTRKESLQSHNVTKAEKPARVSRE 629

Qy        466 LIGDNDNAKKVYEEVLSV-------TPNDGFAKVH------YGFILKAQNKIAESIPYLK 512
Db        630 DLEDVSASKKAKEEAEDVSSTKRQKSPISFFQCVYLDGYNGYGFQFPV-------TPAYR 682

Qy        513 EGIESGDPGT 522
Db        683 PGESSGQPSS 692
```

```
RESULT 4
MDN1_GIALA
ID   MDN1_GIALA     STANDARD;      PRT;  4835 AA.
AC   Q8T5T1;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Midasin (MIDAS-containing protein).
GN   MDN1.
OS   Giardia lamblia (Giardia intestinalis).
OC   Eukaryota; Diplomonadida; Hexamitidae; Giardiinae; Giardia.
OX   NCBI_TaxID=5741;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20389616; PubMed=10930750;
RA   McArthur A.G., Morrison H.G., Nixon J.E., Passamaneck N.Q., Kim U.,
RA   Hinkle G., Crocker M.K., Holder M.E., Farr R., Reich C.I., Olsen G.E.,
RA   Aley S.B., Adam R.D., Gillin F.D., Sogin M.L.;
RT   "The Giardia genome project database.";
RL   FEMS Microbiol. Lett. 189:271-273(2000).
RN   [2]
RP   IDENTIFICATION, GENE NAME, AND SIMILARITY WITH OTHER FAMILY MEMBERS.
RX   PubMed=12102729;
RA   Garbarino J.E., Gibbons I.R.;
RT   "Expression and genomic analysis of midasin, a novel and highly
RT   conserved AAA protein distantly related to dynein.";
RL   BMC Genomics 3:18-18(2002).
CC   -!- FUNCTION: May function as a nuclear chaperone and be involved in
CC       the assembly/disassembly of macromolecular complexes in the
CC       nucleus.
CC   -!- SUBCELLULAR LOCATION: Nuclear (By similarity).
CC   -!- SIMILARITY: Contains 1 VWFA domain.
```

```
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AF494287; AAM12656.1; -.
DR   InterPro; IPR003593; AAA_ATPase.
DR   InterPro; IPR003959; AAA_ATPase_centr.
DR   InterPro; IPR002035; VWF_A.
DR   Pfam; PF00004; AAA; 2.
DR   SMART; SM00382; AAA; 5.
DR   SMART; SM00327; VWA; 1.
DR   PROSITE; PS50234; VWFA; 1.
KW   Chaperone; ATP-binding; Repeat; Nuclear protein.
FT   NP_BIND     356    363       ATP (POTENTIAL).
FT   NP_BIND     814    821       ATP (POTENTIAL).
FT   NP_BIND    1127   1134       ATP (POTENTIAL).
FT   NP_BIND    1513   1520       ATP (POTENTIAL).
FT   NP_BIND    1839   1846       ATP (POTENTIAL).
FT   NP_BIND    3277   3284       ATP (POTENTIAL).
FT   DOMAIN      110    113       POLY-LEU.
FT   DOMAIN     3004   3007       POLY-LEU.
FT   DOMAIN     4156   4161       POLY-ASN.
FT   DOMAIN     4629   4818       VWFA.
SQ   SEQUENCE   4835 AA;  539726 MW;  3A9E12417DB04A50 CRC64;
```

Query Match 4.8%; Score 194; DB 1; Length 4835;
Best Local Similarity 23.1%; Pred. No. 0.084;
Matches 119; Conservative 74; Mismatches 203; Indels 120; Gaps 25;

```
Qy         80 YEEVLGKLGIYD--ADGDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVE 137
Db       4188 HEEQADATGSTDEQAQEDDYNDLDD-KNLSG-----QSDLSVPEADGEDETVNEE---LE 4238

Qy        138 AEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFE 197
Db       4239 EEQQQMSDLSNPD-----QDACAIEEDDDRDLPSSD----ENAEEHDEHEAPVDIDDN-E 4288

Qy        198 TLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQE---------NPDSSEPVVEDER- 247
Db       4289 ASDEQSTYNDNDRDDAINISAQQQATNDEEEMQKDTEYDQENITDSNPDANEVGTNDQKQ 4348

Qy        248 LHHDTDDVTYQVYEEQAVYEPLEN---EGIEITEVTAPPEDNPVEDSQVI-VEEVSIFPV 303
Db       4349 THEDNDQFRQENIEDQWEAESTENSQGEGAESADLKEGNPDMSLEEFQRIWKERLNIHDR 4408

Qy        304 E-------EQQEVPPETNRKTDDPEQKA-------KVKKKKPKLLNKFDKTIKAELDAAE 349
Db       4409 ESEKDEAAEPQDMPLQSNKTVEFDDSKSGRDGALGLTESKHRNLTNQ-------EFDNPN 4461

Qy        350 KLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAE----------------KRR 393
Db       4462 EER---NVEH--NSSCETSQSSHDRPPAEHLNPEISDEGEESSTASDKQEQAVLSHMRES 4516

Qy        394 SNEVLRGAIETYQEVA-SLPD-----VPADL---------LKLSLKRRSDRQQFLGHMRG 438
Db       4517 SKDLLNPEGEVYQELAVSLASEETKRAPEDVAAASARGNHLLLDLIKQTSAAAF------ 4570

Qy        439 SLLTLQRLVQLFPNDTS-LKNDLGVGYLL-------------------IGDNDNAKKVYE 478
Db       4571 SLAERLRII-LEPTVTSDLKGDFRTGKKLNLRRIIPFIASEFQKDKIWLRRTKPSKRVYQ 4629

Qy        479 EVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEG 514
Db       4630 VLLAVDDSSSMAPI-AKYALQAITLLFNACKFLEVG 4664
```

```
RESULT 5
ECX1_METMA
```

```
ID   ECX1_METMA     STANDARD;      PRT;   493 AA.
AC   Q8PTT8;
DT   10-OCT-2003 (Rel. 42, Created)
DT   10-OCT-2003 (Rel. 42, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Probable exosome complex exonuclease 1 (EC 3.1.13.-).
GN   MM2623.
OS   Methanosarcina mazei (Methanosarcina frisia).
OC   Archaea; Euryarchaeota; Methanomicrobia; Methanosarcinales;
OC   Methanosarcinaceae; Methanosarcina.
OX   NCBI_TaxID=2209;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Goe1 / Go1 / ATCC BAA-199 / DSM 3647 / OCM 88;
RX   MEDLINE=22120827; PubMed=12125824;
RA   Deppenmeier U., Johann A., Hartsch T., Merkl R., Schmitz R.A.,
RA   Martinez-Arias R., Henne A., Wiezer A., Baeumer S., Jacobi C.,
RA   Brueggemann H., Lienard T., Christmann A., Boemecke M., Steckel S.,
RA   Bhattacharyya A., Lykidis A., Overbeek R., Klenk H.-P., Gunsalus R.P.,
RA   Fritz H.-J., Gottschalk G.;
RT   "The genome of Methanosarcina mazei: evidence for lateral gene
RT   transfer between Bacteria and Archaea.";
RL   J. Mol. Microbiol. Biotechnol. 4:453-461(2002).
CC   -!- FUNCTION: Probably involved in the 3'->5' degradation of a variety
CC       of RNA species (Potential).
CC   -!- SUBUNIT: Component of the archaeal exosome multienzyme
CC       ribonuclease complex (Potential).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (Potential).
CC   -!- SIMILARITY: Belongs to the RNase PH family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; AE013507; AAM32319.1; -.
DR   HAMAP; MF_00591; -; 1.
DR   InterPro; IPR001247; 3_ExoRNase.
DR   Pfam; PF01138; RNase_PH; 1.
DR   Pfam; PF03725; RNase_PH_C; 1.
KW   Exosome; Hydrolase; Nuclease; Exonuclease; Complete proteome.
FT   DOMAIN        1    254       PROBABLE EXOSOME COMPLEX EXONUCLEASE 1.
FT   DOMAIN      255    493       UNKNOWN.
SQ   SEQUENCE   493 AA;  55248 MW;  979A757BEF8DC090 CRC64;

  Query Match             4.8%;  Score 193;  DB 1;  Length 493;
  Best Local Similarity   26.3%;  Pred. No. 0.0056;
  Matches   97;  Conservative   45;  Mismatches  129;  Indels   98;  Gaps   17;

Qy          61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDV---DDAKVLL---------- 107
               | : |: || |   |  || : ||      |  |: || |    | :: |
Db         153 IPMKGLITSCA---FGKVDGKIVLDLNKEEDNYGEADFPVAMTQDGEITLIQMDGNLTPD 209

Qy         108 ----------GLKERSTSEPAV-------PPEEAEPHTEPE--------EQVPVEAEPQ 141
                        | ||    : ||       | ||    | ||        |  || |  :
Db         210 EIKQGLELVKKGCKEILEIQQAVLRKKFETPVEEVSEETAPEKGAEKEVLEPSPVAAIVE 269

Qy         142 NIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQED--DEFLMAT--DVDDRFE 197
                 :||:|       |:  :| || | |   |  | : |:  :| |  :  |:   |
Db         270 ETPEEAEEP------EVEISEEVEAEILASEVIPDPEDELEEEIEEELEESEEDLETEEE 323

Qy         198 TLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDE---------- 246
               | |   || |    :|| : :|  :::||   | |  :  |  :|:|
Db         324 EFEEEALEEEAEPEEDLEEDLEEDLGEELEEEEEELEEEEFEEEALEEETELEASLECAP 383

Qy         247 ----------RLHHDTDDVTYQVYE------EQAVYEPLENEGIEITEVTAPPEDNPVED 290
                       ||  : :| : :  |       |:|  | || |  || |  |  |:| :|
Db         384 ELKEFDEIEARL--EKEDASIEAEEEIEPEAEEATEEGLEEEA-EIEETAASEEEN-IEA 439

Qy         291 SQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEK 350
                 ||    |  | :|:  |      ::||::   | : |       | :|   :|
Db         440 EAEAEEEAE--PEVEAEEISTEAEEEAEEEPEEE---KSEGPW------KVVKDPSEAG-- 486

Qy         351 LRKRGKIEE 359
                 ||: :|
Db         487 --TRGEKDE 493


RESULT 6
P200_MYCGE
ID   P200_MYCGE     STANDARD;      PRT;  1616 AA.
AC   Q49429; Q49259; Q49298; Q49352; Q49353;
DT   01-NOV-1997 (Rel. 35, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Protein P200.
GN   MG386.
OS   Mycoplasma genitalium.
OC   Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
OX   NCBI_TaxID=2097;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=96026346; PubMed=7569993;
RA   Fraser C.M., Gocayne J.D., White O., Adams M.D., Clayton R.A.,
RA   Fleischmann R.D., Bult C.J., Kerlavage A.R., Sutton G., Kelley J.M.,
RA   Fritchman J.L., Weidman J.F., Small K.V., Sandusky M., Fuhrmann J.L.,
RA   Nguyen D.T., Utterback T.R., Saudek D.M., Phillips C.A., Merrick J.M.,
RA   Tomb J.-F., Dougherty B.A., Bott K.F., Hu P.-C., Lucier T.S.,
RA   Peterson S.N., Smith H.O., Hutchison C.A. III, Venter J.C.;
RT   "The minimal gene complement of Mycoplasma genitalium.";
RL   Science 270:397-403(1995).
RN   [2]
RP   SEQUENCE OF 256-427; 432-543 AND 1083-1140 FROM N.A.
RC   STRAIN=ATCC 33530 / G-37;
RX   MEDLINE=94075230; PubMed=8253680;
RA   Peterson S.N., Hu P.-C., Bott K.F., Hutchison C.A. III;
RT   "A survey of the Mycoplasma genitalium genome by using random
RT   sequencing.";
RL   J. Bacteriol. 175:7918-7930(1993).
CC   -!- FUNCTION: COULD BE AN ACCESSORY STRUCTURAL COMPONENT IN
CC       CYTADHERENCE (BY SIMILARITY).
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U39720; AAC71613.1; -.
DR   EMBL; U02245; AAA03400.1; -.
DR   EMBL; U02245; -; NOT_ANNOTATED_CDS.
DR   EMBL; U02175; AAD12458.1; -.
DR   EMBL; U02126; AAD12402.1; -.
DR   PIR; G64242; G64242.
DR   TIGR; MG386; -.
KW   Cytadherence; Structural protein; Repeat; Complete proteome.
FT   DOMAIN     1205   1389       2 X 32 AA REPEAT.
FT   REPEAT     1205   1236       1-1.
FT   REPEAT     1358   1389       1-2.
FT   DOMAIN      891   1389       2 X 26 AA REPEAT.
FT   REPEAT     1161   1186       2-1.
FT   REPEAT     1310   1339       2-2.
FT   CONFLICT    256    256       P -> S (IN REF. 2).
FT   CONFLICT    304    304       S -> F (IN REF. 2).
SQ   SEQUENCE   1616 AA;  185678 MW;  6AF76A13AC49E4FF CRC64;

  Query Match             4.7%;  Score 189.5;  DB 1;  Length 1616;
  Best Local Similarity   22.3%;  Pred. No. 0.037;
  Matches  142;  Conservative   78;  Mismatches  205;  Indels  213;  Gaps   35;
```

```
Qy        75 FDLVDYEEVLGK----------------LGIYD-------ADGDGDFDVDDAKVL----- 106
             || | :| |  |                    : |          : :  |||   |
Db      1091 FDTVKHEAVFDKNQTQTEGLEEPQVSSEAEVVDQTTTDTVGEPEAVFDVQPEKTTEVKFD 1150

Qy       107 -LGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVE 165
               : :::  ||| |   :   |   ||   :    ::  ::::|   | :    :    : ||
Db      1151 DVENQQKVISEPQVEQQPGEAVFEPSAEAKFDSPVESVQDSQPEPVLEEVQTQPEIQPVE 1210

Qy       166 GEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLE-PEVSHEETEHSYHVEETVSQD-CN 223
               :      |   | :|:|   |          |   ||:| || | |  |:  |||  | |  |
Db      1211 SQPEATFD--TVQPEQTPQE----AKFDSPVETVEQPEFSSEPTQQ--HVESEASFDEPN 1262

Qy       224 QDMEEMMSEQENP--DS----SEPVVEDERLHHDTDDVTYQV--------YEEQAVYEPL 269
               |:|    : :  |  ||    ||| : :  ::| |:  |::        :| |   :|
Db      1263 YDFDEPNYDFDQPSYDSDLQPSEPQYDVDEPNYDFDEPNYEIESKPSEPQFEPQVEQQP- 1321

Qy       270 ENEGIEITEVTAPPE-DNPVE---DSQ--VIVEEVSIFPVEEQQEVPPETNRKTDDPEQK 323
                   |  :  | :|  :  |:|||   |||   ::|||   |  :   |  ||    |  |||
Db      1322 ---GEAVFEPSAEAKFDSPVESVQDSQPEPLLEEVQTQPEIQPVESQPEATFDTVQPEQT 1378

Qy       324 AKVKKKKPKLLNKFDKTKIAEKLDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQ 383
              :         |||                                        ||
Db      1379 PQEA--------KFD-----------------------------------SP-------- 1387

Qy       384 CEDDLAEKRRSNEVLRGAIETYQE--VASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLL 441
                                 :||  ||   |:|  |:|    ::: :  :            ::|
Db      1388 -----------------VETIQEPQVSSEPEV---VVQPNFEERKPE---------TVL 1417

Qy       442 TLQRLVQLFP---NDTSLKNDLGVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYGFIL 498
                :  ::  |       :  ||   :|  ||         :|:     ||          ||    |    ||
Db      1418 EEPQADEIQPEASEEESLDWELLVG-------NNSYGHYEP-------DG-EWVWAGFFG 1462

Qy       499 KAQ--NKIAESIPYLKEGIESGDPGTDDGRFYFHL-GDAMQ-RVGNKEAYKWYELGHKRG 554
                 |  ||  |  :: :  :|               |  |   |  ||  :    |    ||  :  ||      :  |
Db      1463 DDQKWNKDA-TVKWARE------------RDYLPLIGDEVYGRYNNKGEWIWYGFYDESG 1509

Qy       555 HFASV---WQRSLYNVNGLKAQPWWTP----KETGYTELVKSLERNWKLIRDEGLAVMDK 607
               :  |    |:    |    :|    :|        :| || |       | ||
Db      1510 DWVLVDEQWKNRQPRIN--EAPKFWEKLIGNEEYGYYE---DNEWNW------------ 1551

Qy       608 AKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAP 645
                     |        :|:|   |     :       ||:   |    |
Db      1552 -------YDGEFDSEGNWLVFQSEETENLNEDITKDIP 1582
```

RESULT 7
TRDN_RABIT

```
ID   TRDN_RABIT     STANDARD;      PRT;   705 AA.
AC   Q28820; Q28636; Q28637; Q28643;
DT   30-MAY-2000 (Rel. 39, Created)
DT   30-MAY-2000 (Rel. 39, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Triadin.
GN   TRDN.
OS   Oryctolagus cuniculus (Rabbit).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX   NCBI_TaxID=9986;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).
RC   STRAIN=New Zealand white; TISSUE=Skeletal muscle;
RX   MEDLINE=93286104; PubMed=7685347;
RA   Knudson C.M., Stang K.K., Moomaw C.R., Slaughter C.A., Campbell K.P.;
RT   "Primary structure and topological analysis of a skeletal muscle-
RT   specific junctional sarcoplasmic reticulum glycoprotein (triadin).";
RL   J. Biol. Chem. 268:12646-12654(1993).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS SKELETAL).
RC   TISSUE=Skeletal muscle;
RX   MEDLINE=94298946; PubMed=8026576;
RA   Peng M., Fan H., Kirley T.L., Caswell A.H., Schwartz A.;
RT   "Structural diversity of triadin in skeletal muscle and evidence of
RT   its existence in heart.";
RL   FEBS Lett. 348:17-20(1994).
RN   [3]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORMS CARDIAC).
RC   TISSUE=Heart muscle;
RX   MEDLINE=96132942; PubMed=8550602;
RA   Guo W., Jorgensen A.O., Jones L.R., Campbell K.P.;
RT   "Biochemical characterization and molecular cloning of cardiac
RT   triadin.";
RL   J. Biol. Chem. 271:458-465(1996).
RN   [4]
RP   CARBOHYDRATE-LINKAGE SITES, AND INTERCHAIN DISULFIDE BONDS.
RX   MEDLINE=96066664; PubMed=7578102;
RA   Fan H., Brandt N.R., Caswell A.H.;
RT   "Disulfide bonds, N-glycosylation and transmembrane topology of
RT   skeletal muscle triadin.";
RL   Biochemistry 34:14902-14908(1995).
CC   -!- FUNCTION: May be involved in anchoring calsequestrin to the
CC       junctional sarcoplasmic reticulum and allowing its functional
CC       coupling with the ryanodine receptor.
CC   -!- SUBUNIT: Homooligomer of variable subunit number; disulfide-
CC       linked.
CC   -!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC       reticulum.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=6;
CC         Comment=Additional isoforms seem to exist;
CC       Name=Skeletal 1; Synonyms=ST1;
CC         IsoId=Q28820-1; Sequence=Displayed;
CC       Name=Cardiac 1; Synonyms=CT1;
CC         IsoId=Q28820-2; Sequence=VSP_004458, VSP_004460;
CC       Name=Cardiac 2; Synonyms=CT2;
CC         IsoId=Q28820-3; Sequence=VSP_004459, VSP_004461;
CC       Name=Cardiac 3; Synonyms=CT3;
CC         IsoId=Q28820-4; Sequence=VSP_004466;
CC       Name=Skeletal 2; Synonyms=ST2;
CC         IsoId=Q28820-5; Sequence=VSP_004462, VSP_004463, VSP_004464,
CC                                  VSP_004465;
CC       Name=Skeletal 3; Synonyms=ST3;
CC         IsoId=Q28820-6; Sequence=VSP_004464, VSP_004465;
CC   -!- TISSUE SPECIFICITY: Skeletal and cardiac muscle.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; U31540; AAC48496.1; -.
DR   EMBL; L10065; AAA31488.1; -.
DR   EMBL; U31555; AAC48497.1; -.
DR   EMBL; U34201; AAC48498.1; -.
DR   PIR; A45990; A45990.
KW   Transmembrane; Sarcoplasmic reticulum; Glycoprotein;
KW   Alternative splicing.
FT   INIT_MET      0      0
FT   DOMAIN        1     46       CYTOPLASMIC.
FT   TRANSMEM     47     67       POTENTIAL.
FT   DOMAIN       68    705       LUMENAL.
FT   CARBOHYD     74     74       N-LINKED (GLCNAC. . .) (POTENTIAL).
FT   CARBOHYD    624    624       N-LINKED (GLCNAC. . .).
FT   VARSPLIC    264    285       DQYAFCRYMIDIFVHGDLKPGQ -> GGKQSEEAAGCFKRT
FT                                LGKKQMQ (in isoform Cardiac 1).
FT                                /FTId=VSP_004458.
FT   VARSPLIC    264    307       DQYAFCRYMIDIFVHGDLKPGQSPAIPPPSPTEQASRPTPA
FT                                LPT -> ECIFLSAATPQGIPNRQQLNDIHHCFLKTKKGGN
FT                                GQHAFCLKGC (in isoform Cardiac 2).
FT                                /FTId=VSP_004459.
FT   VARSPLIC    286    705       Missing (in isoform Cardiac 1).
FT                                /FTId=VSP_004460.
FT   VARSPLIC    308    705       Missing (in isoform Cardiac 2).
FT                                /FTId=VSP_004461.
```

```
FT   VARSPLIC    415    415       A -> E (in isoform Skeletal 2).
FT                                /FTId=VSP_004462.
FT   VARSPLIC    416    424       Missing (in isoform Skeletal 2).
FT                                /FTId=VSP_004463.
FT   VARSPLIC    584    584       D -> E (in isoform Skeletal 2 and isoform
FT                                Skeletal 3).
FT                                /FTId=VSP_004464.
FT   VARSPLIC    585    591       Missing (in isoform Skeletal 2 and
FT                                isoform Skeletal 3).
FT                                /FTId=VSP_004465.
FT   VARSPLIC    645    705       SKKAKEEAEEVSSTKKQKSPISFFQCVYLDGYNGYGFQFPV
FT                                TPAQYPGESSGKPNSPGPKQ -> LLATVGIWGMNQWMEDL
FT                                SVTLPSK (in isoform Cardiac 3).
FT                                /FTId=VSP_004466.
SQ   SEQUENCE   705 AA;  79003 MW;  13AF1D84475A1361 CRC64;

  Query Match             4.6%;  Score 187;  DB 1;  Length 705;
  Best Local Similarity   21.0%;  Pred. No. 0.018;
  Matches  105;  Conservative   77;  Mismatches  165;  Indels  152;  Gaps   22;

Qy          6 NAKSSGNSSSSGSGSGSTSAGS--------------SSPGARRETKHGGHKNGRKGGLSG 51
              || ::      | :||    | |                |||  |
Db          8 NASTTTTVIDSKNGSVPKSPGKVLKRTVTEDLVTTFSSPAA------------------- 48

Qy         52 TSFFTWFMVIALLGVWTSVAVVWFDLVDYEEV----LGKLG------IYDA--------- 92
                | :||||:  |::|||| ||||||:      : |:|      ::||
Db         49 -----WLLVIALIITWSAVAVVMFDLVDYKNFSASSIAKMGSDPLKLVHDAVEETTDWIY 103

Qy         93 ------------DGDGDFDVDDAKVLLG-LKERSTSEPAVPPEEAEPHTEPEEQVPV--- 136
                        ||| : |  |      | ::|       :  |: |   :|| ::|
Db        104 GFFSLLSDIISSDGDEEDDEGDEDTAKGEIEEPPLKRKDIHKEKIEKQEKPERKIPTKVV 163

Qy        137 ----EAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDV 192
                  | | : :::: | :      :  | | :| ::  :    | | ::   :
Db        164 HKEKEKEKEKVKEKEKPE-----KKATHKEKLEKKEKPETKTVTKEEKKARTK------- 211

Qy        193 DDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDT 252
                | :| :   |       |:    :   | : :   |: |          |:   :|
Db        212 ----EKIEEKTKKE-------VKGVKQEKVKQTVAKAKEVQKTPKPK------EKESKET 254

Qy        253 DDVTYQVYEEQAVY-----EPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQ 307
              |: |  ::|  :    :  ::  :  | |  :| | :                |  |::
Db        255 AAVSKQEQKDQYAFCRYMIDIFVHGDLKPGQSPAIPPPSPTEQAS---RPTPALPTPEEK 311

Qy        308 E---------VPPETNRKTDDPEQKAK--------VKKKKP-KLLNKFDKTIKAELDAAE 349
              |         |  || :| :  : | |        :|||:| |  :    |:|    ||
Db        312 EGEKKKAEKKVTTETKKKAEKEDAKKKSEKETDIDMKKKEPGKSPDTKPGTVKVTTQAAT 371

Qy        350 KLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVA 409
              |  |  | |:  | |:   : |:      || |       ||:  :|    |  | :| |
Db        372 K--KDEKKEDSKKA-KKPAEEQPK------GKKQ------EKKEKHE--EPAKSTKKEHA 414

Qy        410 SLPDVPADLLKLSLKRRSD 428
              :  :   |   ::|: :
Db        415 APSEKQA---KAKIERKEE 430


RESULT 8
CALD_CHICK
ID   CALD_CHICK     STANDARD;      PRT;   771 AA.
AC   P12957; Q03698; Q90756; Q90761; Q92018; Q99230;
DT   01-JAN-1990 (Rel. 13, Created)
DT   01-NOV-1997 (Rel. 35, Last sequence update)
DT   15-MAR-2004 (Rel. 43, Last annotation update)
DE   Caldesmon (CDM).
GN   CALD1 OR CAD.
OS   Gallus gallus (Chicken).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC   Gallus.
OX   NCBI_TaxID=9031;
RN   [1]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).
RC   TISSUE=Gizzard, and Oviduct;
RX   MEDLINE=89340480; PubMed=2760048;
RA   Bryan J., Imai M., Lee R., Moore P., Cook R.G., Lin W.-G.;
RT   "Cloning and expression of a smooth muscle caldesmon.";
RL   J. Biol. Chem. 264:13873-13879(1989).
RN   [2]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM GIZZARD H-CAD).
RC   TISSUE=Gizzard;
RX   MEDLINE=90026426; PubMed=2803315;
RA   Hayashi K., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT   "Primary structure and functional expression of h-caldesmon
RT   complementary DNA.";
RL   Biochem. Biophys. Res. Commun. 164:503-511(1989).
RN   [3]
RP   SEQUENCE FROM N.A., AND PARTIAL SEQUENCE (ISOFORM BRAIN L-CAD).
RC   TISSUE=Brain;
RX   MEDLINE=91093148; PubMed=1824698;
RA   Hayashi K., Fujio Y., Kato I., Sobue K.;
RT   "Structural and functional relationships between h- and
RT   l-caldesmons.";
RL   J. Biol. Chem. 266:355-361(1991).
RN   [4]
RP   SEQUENCE FROM N.A. (GIZZARD H-CAD; BRAIN L-CAD AND GIZZARD L-CAD).
RX   MEDLINE=94071934; PubMed=8250919;
RA   Haruna M., Hayashi K., Yano H., Takeuchi O., Sobue K.;
RT   "Common structural and expressional properties of vertebrate
RT   caldesmon genes.";
RL   Biochem. Biophys. Res. Commun. 197:145-153(1993).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM GIZZARD L-CAD).
RC   TISSUE=Gizzard;
RX   MEDLINE=92042686; PubMed=1939602;
RA   Bryan J., Lee R.;
RT   "Sequence of an avian non-muscle caldesmon.";
RL   J. Muscle Res. Cell Motil. 12:372-375(1991).
RN   [6]
RP   SEQUENCE OF 1-15 FROM N.A. (ISOFORM BRAIN L-CAD).
RC   TISSUE=Brain;
RX   MEDLINE=94271210; PubMed=8002994;
RA   Yano H., Hayashi K., Haruna M., Sobue K.;
RT   "Identification of two distinct promoters in the chicken caldesmon
RT   gene.";
RL   Biochem. Biophys. Res. Commun. 201:618-626(1994).
RN   [7]
RP   SEQUENCE OF 451-756 FROM N.A., AND PARTIAL SEQUENCE.
RC   TISSUE=Gizzard;
RX   MEDLINE=89273666; PubMed=2730665;
RA   Hayashi K., Yamada S., Kanda K., Kimizuka F., Kato I., Sobue K.;
RT   "35 kDa fragment of h-caldesmon conserves two consensus sequences of
RT   the tropomyosin-binding domain in troponin T.";
RL   Biochem. Biophys. Res. Commun. 161:38-45(1989).
RN   [8]
RP   SEQUENCE OF 498-525.
RX   MEDLINE=88293484; PubMed=3401222;
RA   Mornet D., Audemard E., Derancourt J.;
RT   "Identification of a 15 kilodalton actin binding region on gizzard
RT   caldesmon probed by chemical cross-linking.";
RL   Biochem. Biophys. Res. Commun. 154:564-571(1988).
RN   [9]
RP   PHOSPHORYLATION SITES.
RX   MEDLINE=92041815; PubMed=1939059;
RA   Mak A.S., Carpenter M., Smillie L.B., Wang J.H.;
RT   "Phosphorylation of caldesmon by p34cdc2 kinase. Identification of
RT   phosphorylation sites.";
RL   J. Biol. Chem. 266:19971-19975(1991).
RN   [10]
RP   PHOSPHORYLATION OF TYR-27 AND TYR-165.
RX   MEDLINE=20026923; PubMed=10559276;
RA   Wang Z., Danielsen A.J., Maihle N.J., McManus M.J.;
RT   "Tyrosine phosphorylation of caldesmon is required for binding to the
RT   Shc.Grb2 complex.";
RL   J. Biol. Chem. 274:33807-33813(1999).
CC   -!- FUNCTION: Actin- and myosin-binding protein implicated in the
```

```
CC       regulation of actomyosin interactions in smooth muscle and
CC       nonmuscle cells (could act as a bridge between myosin and actin
CC       filaments). Stimulates actin binding of tropomyosin which
CC       increases the stabilization of actin filament structure. In muscle
CC       tissues, inhibits the actomyosin ATPase by binding to F-actin.
CC       This inhibition is attenuated by calcium-calmodulin and is
CC       potentiated by tropomyosin. Interacts with actin, myosin, two
CC       molecules of tropomyosin and with calmodulin. Also play an
CC       essential role during cellular mitosis and receptor capping.
CC   -!- SUBCELLULAR LOCATION: On thin filaments in smooth muscle and on
CC       stress fibers in fibroblasts (nonmuscle) (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=Gizzard h-cad;
CC         IsoId=P12957-1; Sequence=Displayed;
CC       Name=Brain l-cad;
CC         IsoId=P12957-2; Sequence=VSP_004152, VSP_004153;
CC       Name=Gizzard l-cad;
CC         IsoId=P12957-3; Sequence=VSP_004153;
CC   -!- TISSUE SPECIFICITY: High-molecular-weight caldesmon (h-caldesmon)
CC       is predominantly expressed in smooth muscles, whereas low-
CC       molecular-weight caldesmon (l-caldesmon) is widely distributed in
CC       non-muscle tissues and cells. Not expressed in skeletal muscle or
CC       heart.
CC   -!- DOMAIN: The N-terminal part seems to be a myosin/calmodulin-
CC       binding domain, and the C-terminal a tropomyosin/actin/calmodulin-
CC       binding domain. These two domains are separated by a central
CC       helical region in the muscle forms.
CC   -!- PTM: Phosphorylated in non-muscle cells. Phosphorylation by Cdc2
CC       during mitosis causes caldesmon to dissociate from microfilaments.
CC       Phosphorylation reduces caldesmon binding to actin, myosin, and
CC       calmodulin as well as its inhibition of actomyosin ATPase
CC       activity. Phosphorylation also occurs in both quiescent and
CC       dividing smooth muscle cells with similar effects on the
CC       interaction with actin and calmodulin and on microfilaments
CC       reorganization (By similarity).
CC   -!- SIMILARITY: Belongs to the caldesmon family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; J04968; AAA49067.1; -.
DR   EMBL; D17648; BAA04539.1; -.
DR   EMBL; D17634; BAA04539.1; JOINED.
DR   EMBL; D17636; BAA04539.1; JOINED.
DR   EMBL; D17637; BAA04539.1; JOINED.
DR   EMBL; D17638; BAA04539.1; JOINED.
DR   EMBL; D17639; BAA04539.1; JOINED.
DR   EMBL; D17640; BAA04539.1; JOINED.
DR   EMBL; D17641; BAA04539.1; JOINED.
DR   EMBL; D17642; BAA04539.1; JOINED.
DR   EMBL; D17643; BAA04539.1; JOINED.
DR   EMBL; D17644; BAA04539.1; JOINED.
DR   EMBL; D17645; BAA04539.1; JOINED.
DR   EMBL; D17646; BAA04539.1; JOINED.
DR   EMBL; M28417; AAA48810.1; -.
DR   EMBL; M60620; AAA48936.1; -.
DR   EMBL; D17648; BAA04538.1; -.
DR   EMBL; D17635; BAA04538.1; JOINED.
DR   EMBL; D17636; BAA04538.1; JOINED.
DR   EMBL; D17637; BAA04538.1; JOINED.
DR   EMBL; D17639; BAA04538.1; JOINED.
DR   EMBL; D17640; BAA04538.1; JOINED.
DR   EMBL; D17641; BAA04538.1; JOINED.
DR   EMBL; D17642; BAA04538.1; JOINED.
DR   EMBL; D17643; BAA04538.1; JOINED.
DR   EMBL; D17644; BAA04538.1; JOINED.
DR   EMBL; D17645; BAA04538.1; JOINED.
DR   EMBL; D17646; BAA04538.1; JOINED.
DR   EMBL; D17647; BAA04538.1; JOINED.
DR   EMBL; D17648; BAA04540.1; -.
DR   EMBL; D17634; BAA04540.1; JOINED.
DR   EMBL; D17636; BAA04540.1; JOINED.
DR   EMBL; D17637; BAA04540.1; JOINED.
DR   EMBL; D17639; BAA04540.1; JOINED.
DR   EMBL; D17640; BAA04540.1; JOINED.
DR   EMBL; D17641; BAA04540.1; JOINED.
DR   EMBL; D17642; BAA04540.1; JOINED.
DR   EMBL; D17643; BAA04540.1; JOINED.
DR   EMBL; D17644; BAA04540.1; JOINED.
DR   EMBL; D17645; BAA04540.1; JOINED.
DR   EMBL; D17646; BAA04540.1; JOINED.
DR   EMBL; D17647; BAA04540.1; JOINED.
DR   EMBL; M59762; AAA48649.1; -.
DR   EMBL; D17552; BAA04490.1; -.
DR   EMBL; M26684; AAA48811.1; -.
DR   PIR; A33430; A33430.
DR   HSSP; P05412; 1FOS.
DR   InterPro; IPR006017; Caldesmon.
DR   InterPro; IPR006018; Caldesmon_LSP.
DR   Pfam; PF02029; Caldesmon; 1.
DR   PRINTS; PR01076; CALDESMON.
KW   Muscle protein; Actin-binding; Calmodulin-binding; Phosphorylation;
KW   Repeat; Alternative splicing.
FT   DOMAIN       26    199       MYOSIN AND CALMODULIN-BINDING.
FT   DOMAIN      251    390       10 X 13 AA APPROXIMATE TANDEM REPEATS.
FT   REPEAT      251    265       1.
FT   REPEAT      266    278       2.
FT   REPEAT      279    291       3.
FT   REPEAT      294    306       4.
FT   REPEAT      309    321       5.
FT   REPEAT      324    336       6.
FT   REPEAT      337    349       7.
FT   REPEAT      350    362       8.
FT   REPEAT      363    375       9.
FT   REPEAT      378    390       10.
FT   DOMAIN      523    580       TROPOMYOSIN-BINDING (POTENTIAL).
FT   DOMAIN      622    632       TROPOMYOSIN-BINDING (POTENTIAL).

  Query Match          4.5%;  Score 181;  DB 1;  Length 771;
  Best Local Similarity  22.8%;  Pred. No. 0.039;
  Matches  150;  Conservative   87;  Mismatches  248;  Indels  174;  Gaps   32;

Qy        108 GLKERSTSE----PAVPPEE------AEPHTEPEEQV-----PVEAE---------PQNI 143
              | ||   ||      || ||          |  |: || |     | ||           |:|
Db        178 GKKEEKDSEEEKPKEVPTEENQVDVAVEKSTDKEEVVETKTLAVNAENDTNAMLEGEQSI 237

Qy        144 EDEAKEQIQSLLHEMVHAEHVEGEDLQ-QEDGPTGEPQQEDDEFLMATDVDDRFETLEPE 202
              | | :: :    |    |  | | |: :|:    | :|: :|   | :  :| :  | :
Db        238 TDAADKEKEEAEKEREKLEAEEKERLKAEEEKKAAEEKQKAEEEKKAAEERERAKAEEEK 297

Qy        203 VSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPV-VEDERLHHDTDDVTYQVYE 261
              : || | :   ||   :   :: |  :|:|   : |    |:|| :          : |
Db        298 RAAEERERAKAEEE---RKAAEERERAKAEEERKAAEERAKAEEER----------KAAE 344

Qy        262 EQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPETNR------ 315
              |:|  |          |  |  |    |  :   ||      ||::    |  |
Db        345 ERAKAE---------EERKAAEERAKAEKERKAAEERERAKAEEEKRAAEEKARLEAEKL 395

Qy        316 --KTDDPEQKAKVKKKKPKLLNKFDKTIKAELDA-----AEKL----------------- 351
                |    |:||: :| :  || | ::  :|:::|      |||
Db        396 KEKKKMEEKKAQEEKAQANLLRKQEEDKEAKVEAKKESLPEKLQPTSKKDQVKDNKDKEK 455

Qy        352 -----------RKRGKIEE-AVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLR 399
                        ||||  |: | |  :||     :|    :|:: :   |     :| |:
Db        456 APKEEMKSVWDRKRGVPEQKAQNGERELTTPKLKSTENAFGRSNLK-GAANAEAGSEKLK 514

Qy        400 GAIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKND 459
              |  || |    |  | ||   ||| :|:: |                :   |
Db        515 ---EKQQEAA----VELDELK---KRREERRKIL-----------------EEEEQKKK 546
```

```
Qy        460 LGVGYLLIGDNDNAKKVYEEV----------LSVTPNDGFA---KVHYGFILKAQN-KIA 505
Db        547 QEEAERKIREEEEKKRMKEEIERRRAEAAEKRQKVPEDGVSEEKKPFKCFSPKGSSLKIE 606

Qy        506 ESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGNK-EAYKWYELGHKRG---------- 554
Db        607 ERAEFLNKSAQK------SGMKPAHTTAVVSKIDSRLEQYTSAVVGNKAAKPAKPAASDL 660

Qy        555 --------HFASVWQR-SLYNVNGLKAQPWWTPKET-GYTELVKSLERNWKLIRDEGLAV 604
Db        661 PVPAEGVRNIKSMWEKGNVFSSPGGTGTP---NKETAGLKVGVSSRINEWLTKTPEG--- 714

Qy        605 MDKAKGLFLPEDENLREKGDWS-QFTLWQQGRRNENACKGAPKTCTLLEKFPETTGCRR 662
Db        715 -NKSPA---PKPSDLR-PGDVSGKRNLWEKQSVEKPAASSSKVTAT--GKKSETNGLRQ 766
```

RESULT 9
TRDN_HUMAN
ID TRDN_HUMAN STANDARD; PRT; 728 AA.
AC Q13061;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Triadin.
GN TRDN.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Skeletal muscle;
RX MEDLINE=96061957; PubMed=7588753;
RA Taske N.L., Eyre H.J., O'Brien R.O., Sutherland G.R., Denborough M.A.,
RA Foster P.S.;
RT "Molecular cloning of the cDNA encoding human skeletal muscle triadin
RT and its localisation to chromosome 6q22-6q23.";
RL Eur. J. Biochem. 233:258-265(1995).
CC -!- FUNCTION: May be involved in anchoring calsequestrin to the
CC junctional sarcoplasmic reticulum and allowing its functional
CC coupling with the ryanodine receptor (By similarity).
CC -!- SUBUNIT: Homooligomer of variable subunit number; disulfide-linked
CC (By similarity).
CC -!- SUBCELLULAR LOCATION: Type II membrane protein. Sarcoplasmic
CC reticulum.
CC ---
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC ---
DR EMBL; U18985; AAA75315.1; -.
DR PIR; S68191; S68191.
DR Genew; HGNC:12261; TRDN.
DR MIM; 603283; -.
DR GO; GO:0016021; C:integral to membrane; TAS.
DR GO; GO:0006936; P:muscle contraction; TAS.
KW Transmembrane; Sarcoplasmic reticulum; Glycoprotein.
FT INIT_MET 0 0 BY SIMILARITY.
FT DOMAIN 1 46 CYTOPLASMIC.
FT TRANSMEM 47 67 POTENTIAL.
FT DOMAIN 68 728 LUMENAL.
FT DISULFID 269 269 INTERCHAIN (BY SIMILARITY).
FT DISULFID 690 690 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 74 74 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 646 646 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 728 AA; 81423 MW; C1C53BBE1B2A0815 CRC64;

Query Match 4.4%; Score 178.5; DB 1; Length 728;
Best Local Similarity 19.2%; Pred. No. 0.049;
Matches 94; Conservative 86; Mismatches 167; Indels 143; Gaps 20;

```
Qy          6 NAKSSGNSSSSGSGSGSTSAGS--------------SSPGARRETKHGGHKNGRKGGLSG 51
Db          8 NASTTTTVIDSKNGSVPKSPGKVLKRTVTEDIVTTFSSPAA------------------- 48

Qy         52 TSFFTWFMVIALLGVWTSVAVVWFDLVDYEEV----LGKLG-----------------IY 90
Db         49 -----WLLVIALIITWSAVAIVMFDLVDYKNFSASSIAKIGSDPLKLVRDAMEETTDWIY 103

Qy         91 --------------DADGDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPV 136
Db        104 GFFSLLSDIISSEDEEDDDGDEDTDKGEI----DEPPLRKKEIHKDKTEKQEKPERKIQT 159

Qy        137 EAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRF 196
Db        160 KVTHKE-KEKGKEKVRE--KEKPEKKATHKEKIEKKEKPETK-----------TVAKEQK 205

Qy        197 ETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERL----HHDT 252
Db        206 KAKTAEKSEEKTKK--EVKGGKQEKVKQTAAKVKEVQKTPSKPKEKEDKEKAAVSKHEQK 263

Qy        253 DDVTYQVY-------------EEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVS 299
Db        264 DQYAFCRYMIDIFVHGDLKPGQSPAIPPPLPTE-----QASRPTPASPA------LEEKE 312

Qy        300 IFPVEEQQEVPPETNRK-TDDPEQKAK------VKKKKP-KLLNKFDKTIKAELDAA--- 348
Db        313 GEKKKAEKKVTSETKKKEKEDIKKKSEKETAIDVEKKEPGKASETKQGTVKIAAQAAAKK 372

Qy        349 -------EKLRKRGKIEEAVNAFKELVRKY---PQSPRARYGKAQCEDDLAEKRRSNEVL 398
Db        373 DEKKEDSKKTKKPAEVEQPKGKKQEKKEKHVEPAKSPKKEHSVPSDKQVKAKTERAKEEI 432

Qy        399 RGAIETYQEV 408
Db        433 -GAVSSKKAV 441
```

RESULT 10
NFM_RAT
ID NFM_RAT STANDARD; PRT; 845 AA.
AC P12839; Q63370;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Neurofilament triplet M protein (160 kDa neurofilament protein)
DE (Neurofilament medium polypeptide) (NF-M).
GN NEF3 OR NEFM OR NFM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87282618; PubMed=2441012;
RA Napolitano E.W., Chin S.S.M., Colman D.R., Liem R.K.H.;
RT "Complete amino acid sequence and in vitro expression of rat NF-M,
RT the middle molecular weight neurofilament protein.";
RL J. Neurosci. 7:2590-2599(1987).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Wistar;
RX MEDLINE=92332596; PubMed=1321159;
RA Kelly B.M., Gillespie C.S., Sherman D.L., Brophy P.J.;
RT "Schwann cells of the myelin-forming phenotype express neurofilament
RT protein NF-M.";
RL J. Cell Biol. 118:397-410(1992).
RN [3]
RP PHOSPHORYLATION SITES, AND REVISION TO 500.
RX MEDLINE=92165797; PubMed=1537832;
RA Xu Z.-S., Liu W.-S., Willard M.B.;

```
RT   "Identification of six phosphorylation sites in the COOH-terminal
RT   tail region of the rat neurofilament protein M.";
RL   J. Biol. Chem. 267:4467-4471(1992).
RN   [4]
RP   CARBOHYDRATE-LINKAGE SITES.
RX   MEDLINE=93346421; PubMed=8344946;
RA   Dong D.L.-Y., Xu Z.-S., Chevrier M.R., Cotter R.J., Cleveland D.W.,
RA   Hart G.W.;
RT   "Glycosylation of mammalian neurofilaments. Localization of multiple
RT   O-linked N-acetylglucosamine moieties on neurofilament polypeptides
RT   L and M.";
RL   J. Biol. Chem. 268:16679-16687(1993).
CC   -!- FUNCTION: Neurofilaments usually contain three intermediate
CC       filament proteins: L, M, and H which are involved in the
CC       maintenance of neuronal caliber.
CC   -!- PTM: There are a number of repeats of the tripeptide K-S-P, NFM is
CC       phosphorylated on a number of the serines in this motif. It is
CC       thought that phosphorylation of NFM results in the formation of
CC       interfilament cross bridges that are important in the maintenance
CC       of axonal caliber.
CC   -!- PTM: Phosphorylation seems to play a major role in the functioning
CC       of the larger neurofilament polypeptides (NF-M and NF-H), the
CC       levels of phosphorylation being altered developmentally and
CC       coincident with a change in the neurofilament function.
CC   -!- SIMILARITY: Belongs to the intermediate filament family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; M18628; AAA41696.1; -.
DR   EMBL; Z12152; CAA78136.1; -.
DR   PIR; A45669; A45669.
DR   GlycoSuiteDB; P12839; -.
DR   InterPro; IPR006821; Filament_head.
DR   InterPro; IPR001664; IF.
DR   InterPro; IPR002957; Keratin_I.
DR   Pfam; PF00038; filament; 1.
DR   Pfam; PF04732; filament_head; 1.
DR   PRINTS; PR01248; TYPE1KERATIN.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Neurone; Phosphorylation;
KW   Glycoprotein.
FT   INIT_MET      0      0       BY SIMILARITY.
FT   DOMAIN        1    103       HEAD.
FT   DOMAIN      103    410       ROD.
FT   DOMAIN      411    844       TAIL.
FT   DOMAIN      103    134       COIL 1A.
FT   DOMAIN      135    147       LINKER 1.
FT   DOMAIN      148    246       COIL 1B.
FT   DOMAIN      247    263       LINKER 12.
FT   DOMAIN      264    285       COIL 2A.
FT   DOMAIN      286    289       LINKER 2.
FT   DOMAIN      290    410       COIL 2B.
FT   CARBOHYD     47     47       O-LINKED (GLCNAC).
FT                                /FTId=CAR_000130.
FT   CARBOHYD    430    430       O-LINKED (GLCNAC).
FT                                /FTId=CAR_000131.
FT   MOD_RES     502    502       PHOSPHORYLATION.
FT   MOD_RES     506    506       PHOSPHORYLATION.
FT   MOD_RES     536    536       PHOSPHORYLATION.
FT   MOD_RES     603    603       PHOSPHORYLATION.
FT   MOD_RES     608    608       PHOSPHORYLATION.
FT   MOD_RES     666    666       PHOSPHORYLATION.
FT   CONFLICT     17     17       MISSING (IN REF. 2).
FT   CONFLICT     21     21       R -> P (IN REF. 2).
FT   CONFLICT    204    204       V -> L (IN REF. 2).
FT   CONFLICT    500    500       MISSING (IN REF. 1).
SQ   SEQUENCE   845 AA;  95660 MW;  316C41655B11197D CRC64;
```

```
  Query Match          4.4%;  Score 176.5;  DB 1;  Length 845;
  Best Local Similarity  23.3%;  Pred. No. 0.074;
  Matches   90;  Conservative   59;  Mismatches  137;  Indels  101;  Gaps   15;

Qy        81 EEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEP 140
             ||:: :  : |       :::||  ::  :  ::::     || |   | ||:  ||  |
Db       453 EEIIEETKVEDEKS----EMEDALTVIAEELAASAK-----EEKEEAEEKEEEPEVEKSP 503

Qy       141 QNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLE 200
                ||||:               |||  ::|:|   | ::|:||: :  :   ::     |
Db       504 VK-SPEAKEE------------EEGEKEEEEEGQ--EEEEEEDEGVKSDQAEEGGS--E 545

Qy       201 PEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDTDDVTYQVY 260
               | | |: |                    |||    :|   | |       :  |
Db       546 KEGSSEKDE--------------------GEQEEEGETEAEGEGEEAEAKEEKKTEGKV 584

Qy       261 EEQAVYEPLENEGIE----------ITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVP 310
             || |: | :: |  |          : ||   ||    :| |   |:     {||::||
Db       585 EEMAIKEEIKVEKPEKAKSPVPKSPVEEVKPKPEAKAGKDEQKEEEK-----VEEKKEVA 639

Qy       311 PET--NRKTDDPEQKAK----------VKKKKPKLLNKFDKTIKAELDAAEKLRK---- 353
               |:    | :  |:| |            ||:|  : :    |::|  |:   |  |
Db       640 KESPKEEKVEKKEEKPKDVPDKKKAESPVKEKAVEEMITITKSVKVSLEKDTKEEKPQQQ 699

Qy       354 ---RGKIEEAVNAFKELVRKYPQSPRAR--------YGKAQCEDDLAEKRRSNEVLRGAI 402
                : | ||   : :|:  | ||  :           || : | :  ||    |  :| :
Db       700 EKVKEKAEEEGGSEEEVGDKSPQESKKEDIAINGEVEGKEEEEQETQEKGSGQEEEKGVV 759

Qy       403 ETYQEVASLPDVPADLLKLSLKRRSDR 429
               :|:     ||:      |: ||
Db       760 TNGLDVS-----PAE-----EKKGEDR 776
```

```
RESULT 11
MYS3_HUMAN
ID   MYS3_HUMAN     STANDARD;      PRT;  2004 AA.
AC   Q92794;
DT   15-JUL-1998 (Rel. 36, Created)
DT   15-JUL-1998 (Rel. 36, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   MYST histone acetyltransferase 3 (Runt-related transcription factor
DE   binding protein 2) (Monocytic leukemia zinc finger protein) (Zinc
DE   finger protein 220).
GN   MYST3 OR RUNXBP2 OR ZNF220 OR MOZ.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=96376968; PubMed=8782817;
RA   Borrow J., Stanton V.P. Jr., Andresen J.M., Becher R., Behm F.G.,
RA   Chaganti R.S.K., Civin C.I., Disteche C., Dube I., Frischauf A.M.,
RA   Horsman D., Mitelman F., Volinia S., Watmore A.E., Housman D.E.;
RT   "The translocation t(8;16)(p11;p13) of acute myeloid leukaemia fuses
RT   a putative acetyltransferase to the CREB-binding protein.";
RL   Nat. Genet. 14:33-41(1996).
CC   -!- FUNCTION: May represent a chromatin-associated acetyltransferase.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: Involved in acute myeloid leukemias through a chromosomal
CC       translocation t(8;16)(p11;p13) involving MYST3 and CREBBP.
CC   -!- SIMILARITY: Contains 2 PHD-type zinc fingers.
CC   -!- SIMILARITY: Belongs to the MYST (SAS/MOZ) family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
```

```
DR   EMBL; U47742; AAC50662.1; -.
DR   Genew; HGNC:13013; MYST3.
DR   MIM; 601408; -.
DR   GO; GO:0006323; P:DNA packaging; TAS.
DR   InterPro; IPR005818; Histone_H1/H5.
DR   InterPro; IPR002717; MOZ_SAS.
DR   InterPro; IPR001965; Znf_PHD.
DR   Pfam; PF01853; MOZ_SAS; 1.
DR   Pfam; PF00628; PHD; 2.
DR   SMART; SM00526; H15; 1.
DR   SMART; SM00249; PHD; 2.
DR   PROSITE; PS01359; ZF_PHD_1; 1.
DR   PROSITE; PS50016; ZF_PHD_2; 2.
KW   Proto-oncogene; Chromosomal translocation; Zinc-finger; Repeat;
KW   Nuclear protein.
FT   ZN_FING      206    265       PHD-TYPE 1.
FT   ZN_FING      259    313       PHD-TYPE 2.
FT   DOMAIN       371    379       POLY-SER.
FT   ZN_FING      538    560       C2HC-TYPE.
FT   DOMAIN       788    801       POLY-GLU.
FT   DOMAIN       989    995       POLY-GLU.
FT   DOMAIN      1019   1026       POLY-ARG.
FT   DOMAIN      1069   1078       POLY-GLU.
FT   DOMAIN      1147   1150       POLY-LYS.
FT   DOMAIN      1221   1242       GLU-RICH.
FT   DOMAIN      1267   1302       GLU-RICH.
FT   DOMAIN      1411   1414       POLY-GLU.
FT   DOMAIN      1593   1597       POLY-SER.
FT   DOMAIN      1643   1704       GLN/PRO-RICH.
FT   DOMAIN      1897   1977       MET-RICH.
FT   SITE        1546   1547       BREAKPOINT FOR TRANSLOCATION TO FORM
FT                                 MOZ-CBP.
SQ   SEQUENCE   2004 AA;  225054 MW;  9FFBBAC3792854BA CRC64;

  Query Match             4.4%;  Score 175;  DB 1;  Length 2004;
  Best Local Similarity   22.7%;  Pred. No. 0.26;
  Matches   73;  Conservative   48;  Mismatches  113;  Indels   88;  Gaps   13;

Qy        122 EEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQ 181
Db       1205 QESEETVEPKEDMP-------LPEERKEE-EEMQAEAEEAEEGEEEDAASSEVPAASPAD 1256

Qy        182 EDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEP 241
Db       1257 SSN--------SPETETKEPEVEEEEEKPRVSEEQRQSEEEQQELEEPEPEEEEDAAAET 1308

Qy        242 VVEDERLHHDTDD--------VTYQVYEEQAVYEPLENE-GI------------------ 274
Db       1309 AQNDD---HDADDEDDGHLESTKKKELEEQPTREDVKEEPGVQESFLDANMQKSREKIKD 1365

Qy        275 -EITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKL 333
Db       1366 KEETELDS-EEEQPSHDTSVVSEQMA--------------GSEDDHEEDSHTKEELIEL 1409

Qy        334 LNKFDKTIKAELDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRR 393
Db       1410 KEE-EEIPHSELDL-----------ETVQAVQSLTQEESSEHEGAY--QDCEETLA---- 1451

Qy        394 SNEVLRGAIETYQEVASLPDVP 415
Db       1452 -------ACQTLQSYTQADEDP 1466


RESULT 12
YNJ1_YEAST
ID   YNJ1_YEAST     STANDARD;      PRT;  1240 AA.
AC   P53935;
DT   01-OCT-1996 (Rel. 34, Created)
DT   01-OCT-1996 (Rel. 34, Last sequence update)
DT   01-OCT-1996 (Rel. 34, Last annotation update)
DE   Hypothetical 141.5 kDa protein in YPT53-RHO2 intergenic region.
GN   YNL091W OR N2231.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / FY1679;
RX   MEDLINE=96367601; PubMed=8771715;
RA   Garcia-Cantalejo J.M., Boskovic J., Jimenez A.;
RT   "Sequence analysis of a 14.2 kb fragment of Saccharomyces cerevisiae
RT   chromosome XIV that includes the ypt53, tRNALeu and gsr m2 genes and
RT   four new open reading frames.";
RL   Yeast 12:599-608(1996).
CC   -!- SIMILARITY: TO S.POMBE SPAC29E6.10C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X85811; CAA59826.1; -.
DR   EMBL; Z71367; CAA95967.1; -.
DR   PIR; S52734; S52734.
DR   GermOnline; 143097; -.
DR   SGD; S0005035; YNL091W.
DR   GO; GO:0009651; P:salinity response; IMP.
KW   Hypothetical protein.
FT   DOMAIN       756    761       POLY-GLU.
SQ   SEQUENCE   1240 AA;  141513 MW;  3FE9D265822D5778 CRC64;

  Query Match             4.3%;  Score 172.5;  DB 1;  Length 1240;
  Best Local Similarity   23.9%;  Pred. No. 0.19;
  Matches  103;  Conservative   66;  Mismatches  175;  Indels   87;  Gaps   21;

Qy        144 EDEAKEQIQSLLHEMVHAEHVE-GEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPE 202
Db        460 EDEDEEDYDD------YSEYAEDSEEVSEYEGIEAVEKPEHDE-----KSNGIRETLHLS 508

Qy        203 VSHE-----ETEHSYHVEETVSQDCNQDMEEMMSEQENP----------------DSSEP 241
Db        509 YDHDHKRQNHPHHHYHSTSTHSED-ELSEEEYISDIELPHDPHKHFHRDDDILDGDEDEP 567

Qy        242 VVEDERLHHDTDDVTYQ--VYEEQAVYEPLENEGIEITE-----VTAPPEDNPVEDSQV- 293
Db        568 EEEDENEGDDEED-TYDSGLDETDRLEEGRKLIQIAITKLLQSRIMASYHEKQADNNRLK 626

Qy        294 IVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKPDK--------TIKAEL 345
Db        627 LLQE-----LEEEKRKKREKEEKKQKKREKEKEKKRLQQLAKEEEKRKREEEKERLKKEL 681

Qy        346 DAAEKLRK---RGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRR--SNEVLRG 400
Db        682 EEREMRRREAQRKKVEEA-------KRKKDEERKRRLEEQQRREEMQEKQRKQKEELKRK 734

Qy        401 AIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRL--VQLFPNDTSLK- 457
Db        735 REEEKKRIREQKRLEQEKLQKE-KEEEERQRLIAE---DALRKQKLNEEQTSANILSAKP 790

Qy        458 -NDLGVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIE 516
Db        791 FTENGVGNPVSSQSHPNMTNYQEDNSCSINDE--------ILKMVNSVAASKPVSPTGFN 842

Qy        517 SGD---PGTDD 524
Db        843 VHDLLLPSTNN 853


RESULT 13
YHU6_YEAST
ID   YHU6_YEAST     STANDARD;      PRT;   465 AA.
AC   P38845;
```

```
DT   01-FEB-1995 (Rel. 31, Created)
DT   01-FEB-1995 (Rel. 31, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Hypothetical 51.1 kDa protein in DCD1-MRPL6 intergenic region.
GN   YHR146W.
OS   Saccharomyces cerevisiae (Baker's yeast).
OC   Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC   Saccharomycetales; Saccharomycetaceae; Saccharomyces.
OX   NCBI_TaxID=4932;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=S288c / AB972;
RX   MEDLINE=94378003; PubMed=8091229;
RA   Johnston M., Andrews S., Brinkman R., Cooper J., Ding H., Dover J.,
RA   Du Z., Favello A., Fulton L., Gattung S., Geisel C., Kirsten J.,
RA   Kucaba T., Hillier L.W., Jier M., Johnston L., Langston Y.,
RA   Latreille P., Louis E.J., Macri C., Mardis E., Menezes S., Mouser L.,
RA   Nhan M., Rifkin L., Riles L., St Peter H., Trevaskis E., Vaughan K.,
RA   Vignati D., Wilcox L., Wohldman P., Waterston R., Wilson R.,
RA   Vaudin M.;
RT   "Complete nucleotide sequence of Saccharomyces cerevisiae chromosome
RT   VIII.";
RL   Science 265:2077-2082(1994).
CC   -!- SIMILARITY: TO YEAST YNL173C.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; U10397; AAB68982.1; -.
DR   PIR; S46759; S46759.
DR   GermOnline; 139464; -.
DR   SGD; S0001189; CRP1.
DR   GO; GO:0005634; C:nucleus; IC.
DR   GO; GO:0003677; F:DNA binding; IDA.
KW   Hypothetical protein.
SQ   SEQUENCE   465 AA;  51115 MW;  30880758F37991C7 CRC64;

  Query Match             4.3%;  Score 172;  DB 1;  Length 465;
  Best Local Similarity   22.6%;  Pred. No. 0.06;
  Matches   90;  Conservative   52;  Mismatches  132;  Indels  124;  Gaps   16;

Qy         76 DLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLG---------------------------- 108
              |||: :|| |   | :| |           :| |
Db         92 DLVETQEVAGASRIPEAGG----------LLCGKPPRSAGPPSTSNRKKNKRNNKKRRSK 141

Qy        109 LKERSTSEPAVPPEEAEPHTEPE---------------EQVPVEAEPQNIEDEAKEQIQS 153
              ||::||       |  : : | :               |: |: ||| |:  ||
Db        142 LKKKSTKNNKKSNESLDDNEEEDGVTGTTTEDVTGTSREETPL-AEPTNVSKEAPGNFHI 200

Qy        154 LLHEMVHAEHVEGEDLQQEDGPTGEP-------QQEDDEFLMATDVDDRFETLEPEVSHE 206
              |  :       :  |  | :|  | |          |  ||    ||| |          |
Db        201 LPID-------QSADTTQSNGIIGGPGPVLVPNPGEIKEFTEIRDVDAR----------E 243

Qy        207 ETEHSYHVEETVSQDCNQDMEEMMSEQEN--PDSSEPVVEDERLHHDTDDVTYQVYEEQA 264
                |    ||          :|  ::|:    | : :|:|| : : |:  ::| ||   :|
Db        244 LNERLNKKEEVPEPVAGPIVESSVTEKSPALPQADDPIVETKEVAHNVQELTPQV---EA 300

Qy        265 VYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVS-IFPVEEQQEVPPETNRKTDDPEQK 323
              |  || ||          ||  |  ::|: : | | : |||    :
Db        301 V-TPLINE----------PEPLPTPEAQISIPESSKVEPVEGSLQ--------------- 334

Qy        324 AKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEAVNAFKELVRKYPQSP-----RAR 378
                    ||: | : | :  || ::|:  : | :| |      :| | |: |      |
Db        335 -------SKLVEKREST-EGVLDGSKKVENKAKKDEEVFTLDPIVNKAPKLPLTDEQTAE 386

Qy        379 YGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPA 416
                |:   :  ||::  |  :|: |  :   |    |:
Db        387 GRKSPAVSEEKEKKKKQE--KGSKEVKRSETSKEKKPS 422
```

```
RESULT 14
YCF2_OENVI
ID   YCF2_OENVI     STANDARD;      PRT;   630 AA.
AC   P31569;
DT   01-JUL-1993 (Rel. 26, Created)
DT   01-JUL-1993 (Rel. 26, Last sequence update)
DT   16-OCT-2001 (Rel. 40, Last annotation update)
DE   Hypothetical protein ycf2 (ORF 2280) (Fragment).
GN   YCF2.
OS   Oenothera villaricae.
OG   Chloroplast.
OC   Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
OC   Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; rosids;
OC   Myrtales; Onagraceae; Oenothera.
OX   NCBI_TaxID=3941;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=93169690; PubMed=8435856;
RA   Nimzyk R., Shoendorf T., Hachtel W.;
RT   "In-frame length mutations associated with short tandem repeats are
RT   located in unassigned open reading frames of Oenothera chloroplast
RT   DNA.";
RL   Curr. Genet. 23:265-270(1993).
CC   -!- SIMILARITY: Belongs to the ycf2 family.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X64615; CAA45896.1; -.
DR   PIR; S29796; S29796.
KW   Chloroplast; Hypothetical protein.
FT   NON_TER       1      1
SQ   SEQUENCE   630 AA;  72781 MW;  6AEFFF7DC75B0BAA CRC64;

  Query Match             4.3%;  Score 172;  DB 1;  Length 630;
  Best Local Similarity   20.9%;  Pred. No. 0.087;
  Matches  107;  Conservative   82;  Mismatches  168;  Indels  154;  Gaps   22;

Qy         52 TSFFTWF---------MVIALLGVWTSVAVVWFDLV-----DYEEVLGKLGIYDADGDGD 97
              :| : |:          : | |  :  |   :  ||:     | : ::   |: : | |
Db        100 SSLYKWYFELGTSMKKLTILLYLLTCSAGSIAQDLLSPPGPDEQNLITSYGLVENDSDLV 159

Qy         98 FDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDE---AKEQIQSL 154
                : |  :: || |   :      | | |     ||    | | :  |:|    :|:::
Db        160 HGLSD--IVHGLLELEGALVGSSPTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGT 217

Qy        155 LHEMVHA--EHVEG-----EDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPEV--SH 205
              | |     | |||      |  ::|    | |  :  :| :  |: |:  |  | || :
Db        218 EDEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTE-DEEVEGTEEEVEGTE 276

Qy        206 EETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPVV---EDERLHHDTDDVT------ 256
              || | :    |   ::     ||:   :|  : :|  |   | :     | | ||
Db        277 EEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEEEVEGTEKDSSQFDNDRVTLLLRPK 336

Qy        257 ------------YQVYE---------EQAVYEP---LENEGIEIT-------------- 277
                         :| ||          :: |: |    ||:   |:
Db        337 PRNPLDIQRLIYQHQKYESELEEDDDDDEDVFAPQKMLEDLFSELVWSPRIWHPWDFLLD 396

Qy        278 -EVTAPPEDNPVEDSQV----IVEEVSIFPVEE----------------QQEVPPETNRK 316
               |   | |: | |: ::     :  ||::: |||                 : |:  | : :
Db        397 CEAEIPAEEIPEEEDELPEDALETEVAVWGVEEEGEADDEEDVLLEAQQEDELLEEEDEE 456

Qy        317 TDDPEQKAKVKKKKPK----LLNKFDKTIKAELDAAEKL---------------------- 351
               |: | :   ::::||     |:: ::  : | :  ::|
Db        457 LDEEEDELDEEEEEPKEEEDELHEEEEEEEEEEEEEDELQENDSEFFRVKPIIPRHRWIF 516
```

```
Qy        352 RKRGKI-------EEAVNAFKELV--------RKYPQSPRARYGKAQCEDD--------- 387
              ||:  :       |||    |||:        |  |: || |:   : :|
Db        517 RKKKDVFEVLSYPEEATEISKELLRLLNPKTKRDAPKRPRQRWWTKKKQDKHYELLLDRQ 576

Qy        388 --LAEKR---RSNEVLRG--AIETYQEVASL 411
                |  ||   :||   |     |:|| :::|
Db        577 RWLITKRSLSKSNGFFRSNTPSESYQYLSNL 607


RESULT 15
LAM2_MOUSE
ID   LAM2_MOUSE     STANDARD;      PRT;   592 AA.
AC   P21619;
DT   01-MAY-1991 (Rel. 18, Created)
DT   01-MAY-1991 (Rel. 18, Last sequence update)
DT   10-OCT-2003 (Rel. 42, Last annotation update)
DE   Lamin B2.
GN   LMNB2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=91106216; PubMed=2102682;
RA   Hoeger T.H., Zatloukal K., Waizenegger I., Krohne G.;
RT   "Characterization of a second highly conserved B-type lamin present
RT   in cells previously thought to contain only a single B-type lamin.";
RL   Chromosoma 99:379-390(1990).
RN   [2]
RP   ERRATUM.
RX   MEDLINE=91339548; PubMed=2102440;
RA   Hoeger T.H., Zatloukal K., Waizenegger I., Krohne G.;
RL   Chromosoma 100:67-69(1990).
CC   -!- FUNCTION: Lamins are components of the nuclear lamina, a fibrous
CC       layer on the nucleoplasmic side of the inner nuclear membrane,
CC       which is thought to provide a framework for the nuclear envelope
CC       and may also interact with chromatin.
CC   -!- SUBCELLULAR LOCATION: Nucleoplasmic side of the inner nuclear
CC       membrane.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=B2;
CC         IsoId=P21619-1; Sequence=Displayed;
CC       Name=B3;
CC         IsoId=P48680-1; Sequence=External;
CC         Note=No experimental confirmation available;
CC   -!- PTM: B-TYPE LAMINS UNDERGO A SERIES OF MODIFICATIONS, SUCH AS
CC       FARNESYLATION AND PHOSPHORYLATION. INCREASED PHOSPHORYLATION OF
CC       THE LAMINS OCCURS BEFORE ENVELOPE DISINTEGRATION AND PROBABLY
CC       PLAYS A ROLE IN REGULATING LAMIN ASSOCIATIONS.
CC   -!- MISCELLANEOUS: The structural integrity of the lamina is strictly
CC       controlled by the cell cycle, as seen by the disintegration and
CC       formation of the nuclear envelope in prophase and telophase,
CC       respectively.
CC   -!- SIMILARITY: Belongs to the intermediate filament family. THIS IS A
CC       B TYPE LAMIN.
CC   ---------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------
DR   EMBL; X54098; CAA38032.1; -.
DR   PIR; B48315; B48315.
DR   MGD; MGI:96796; Lmnb2.
DR   GO; GO:0005638; C:lamin filament; IDA.
DR   InterPro; IPR001664; IF.
DR   InterPro; IPR001322; IF_tail_C.
DR   Pfam; PF00038; filament; 1.
DR   Pfam; PF00932; IF_tail; 1.
DR   PROSITE; PS00226; IF; 1.
KW   Intermediate filament; Coiled coil; Nuclear protein; Lipoprotein;
KW   Prenylation; Phosphorylation; Alternative splicing.
FT   DOMAIN       1    26       HEAD.
FT   DOMAIN      27   378       ROD.
FT   DOMAIN     379   592       TAIL.
FT   DOMAIN      27    61       COIL 1A.
FT   DOMAIN      62    73       LINKER 1.
FT   DOMAIN      74   207       COIL 1B.
FT   DOMAIN     208   234       LINKER 2.
FT   DOMAIN     235   378       COIL 2.
FT   DOMAIN     414   419       NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT   DOMAIN     561   574       ASP/GLU-RICH
FT                              (HIGHLY ACIDIC, COULD BE INVOLVED IN
FT                              CHROMATIN BINDING).
FT   LIPID      589   589       S-farnesyl cysteine (By similarity).
SQ   SEQUENCE   592 AA;  67029 MW;  7D5AC51BC9A5041E CRC64;

  Query Match             4.2%;  Score 169.5;  DB 1;  Length 592;
  Best Local Similarity   20.6%;  Pred. No. 0.11;
  Matches  129;  Conservative  113;  Mismatches  244;  Indels  141;  Gaps   29;

Qy         78 VDYEEVLGKLGIYDADGDGDFDVDDAKVLLG--LKERSTSEPAVPPEEAEPHTEPEEQVP 135
              |   || |   :|::      :: ||: :|    :||:  :  :   :||     :
Db         62 VTTREVSGIKTLYES------ELADARRVLDETARERARLQIEIGKVQAELEEARKSAKK 115

Qy        136 VEAEPQNIEDEAKEQIQSLLH----EMVHA----EHVEGEDLQ-------QEDGPTGEPQ 180
              | |    :   |: ::|| |     |:  |    | :| |  :         |||     :
Db        116 REGELTVAQGRVKD-LESLFHRSEAELATALSDNEGLETEVAELRAQLAKAEDGHAVAKK 174

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPD-SS 239
              | : | ||  |:::| ::|: |::  ::      | ||  :   : :|   | |:  |
Db        175 QLEKETLMRVDLENRCQSLQEELAFSKSVPEEEVRETRRRHERRLVEVDSSRQQEYDFKM 234

Qy        240 EPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVS 299
              :|| |  ||      |:|  ||      |:|  :
Db        235 AQALEDLRSQHDEQVRLYRVELEQTYQAKLDNAKL------------------------ 269

Qy        300 IFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEE 359
              : :| :      |: :  | : :|:    :||      :: :  |||      ::||
Db        270 ---LSDQNDKAAHAARE-ELKEARMRVESLSYQLLG-----LQKQASAAE--NHIHELEE 318

Qy        360 AVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLP---DVPA 416
              |:   ::  ||   :          | :: | |   : ::  :  |||: :  :    |:
Db        319 ALRGERDKFRKMLDAK---------EQEMTEVR---DRMQQQLAEYQELLDIKLALDMEI 366

Qy        417 DLLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKV 476
              :  |:  :| : |      | :|: |          ||  :  ||| : :|     ::
Db        367 SAYRKLLEGEEERLK-LSPSPSSRITISRA-------TSSSSSSGVG-MSVGQRGGKRRR 417

Qy        477 YEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRF--YFHLGDA 534
              |:  | :|:   ::|  |      :::|:::      |:  ||   :|||    :  |
Db        418 LEDT-SGSPSRA-SRVSSG------SRLAQTVATGVVNIDEVDP---EGRFVRLKNSSDK 466

Qy        535 MQRVGNKE----------AYKW---YELGHKRGHFASVWQRSLYNVNG------LKAQP 574
              | :||            |||:   | |  : |   :||       :       |:|
Db        467 DQSLGNWRIKRQVLEGEDIAYKFTPKYVL--RAGQTVTVWAAGAGATHSPPSTLVWKSQT 524

Qy        575 WWTPKETGYTELVKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQG 634
              | | |:  | || :       :   :  :|  :  |     :|   |: :: :  |: |
Db        525 NWGPGESFRTALVSADGEEVAVKAAKHSSVQGRENG-----EEEEEEEAEFGEEDLFHQ- 578

Qy        635 RRNENACKGAPKTCTLLEKFPETTGCR 661
                    :| |:|         : |||
Db        579 -------QGDPRT--------TSRGCR 590


Search completed: May  5, 2004, 11:02:51
Job time : 20 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 5, 2004, 10:59:38 ; Search time 52 Seconds
(without alignments)
4599.285 Million cell updates/sec

Title: US-09-903-216-2
Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSSGSGS.....IVDVWHPELTPQQRRSLPAI 758

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4022	100.0	758	4	Q9Y4J0	Q9y4j0 homo sapien
2	3209.5	79.8	739	11	Q9EPA6	Q9epa6 mus musculu
3	3208.5	79.8	741	11	Q8BSY0	Q8bsy0 mus musculu
4	3175.5	79.0	725	11	Q8BQK0	Q8bqk0 mus musculu
5	2950	73.3	658	11	Q8CBM2	Q8cbm2 mus musculu
6	2919.5	72.6	689	11	Q9EQ66	Q9eq66 mus musculu
7	1636	40.7	313	4	Q9H2C4	Q9h2c4 homo sapien
8	1457.5	36.2	299	4	Q9H291	Q9h291 homo sapien
9	1376.5	34.2	270	4	Q8TB28	Q8tb28 homo sapien
10	1066.5	26.5	785	5	Q9GQ82	Q9gq82 drosophila
11	998.5	24.8	308	11	Q9EQ65	Q9eq65 mus musculu
12	972.5	24.2	292	11	Q91WG6	Q91wg6 mus musculu
13	890	22.1	270	11	Q920F9	Q920f9 mus musculu
14	798	19.8	872	5	Q93178	Q93178 caenorhabdi
15	779.5	19.4	259	11	Q920F8	Q920f8 mus musculu
16	708.5	17.6	258	11	Q9EQ67	Q9eq67 mus musculu
17	691.5	17.2	215	11	Q920F7	Q920f7 mus musculu
18	550.5	13.7	133	11	Q9EQ63	Q9eq63 mus musculu
19	533.5	13.3	147	11	Q9EQ62	Q9eq62 mus musculu
20	516	12.8	245	11	Q8CH79	Q8ch79 mus musculu
21	435.5	10.8	210	4	Q9NRI1	Q9nri1 homo sapien
22	420.5	10.5	210	6	Q28264	Q28264 canis famil
23	418	10.4	225	4	Q9NRI0	Q9nri0 homo sapien
24	407.5	10.1	207	11	Q9EQ64	Q9eq64 mus musculu
25	407.5	10.1	207	11	Q9CR06	Q9cr06 mus musculu
26	407.5	10.1	212	11	Q9D7J8	Q9d7j8 mus musculu
27	390.5	9.7	304	2	Q53792	Q53792 streptomyce
28	276	6.9	343	11	Q80VP9	Q80vp9 mus musculu
29	275	6.8	250	16	Q93H17	Q93h17 streptomyce
30	274.5	6.8	343	4	Q9UH39	Q9uh39 homo sapien
31	263.5	6.6	186	4	Q9NSN3	Q9nsn3 homo sapien
32	257	6.4	176	6	Q9N1E7	Q9n1e7 oryctolagus
33	251	6.2	176	11	Q9CUZ2	Q9cuz2 mus musculu
34	250	6.2	49	11	Q9EQ69	Q9eq69 mus musculu
35	217.5	5.4	230	4	Q8N4H3	Q8n4h3 homo sapien
36	217.5	5.4	250	4	Q8N316	Q8n316 homo sapien
37	217.5	5.4	280	4	Q8IW63	Q8iw63 homo sapien
38	217.5	5.4	347	4	Q96H00	Q96h00 homo sapien
39	216.5	5.4	1110	13	Q91255	Q91255 petromyzon
40	207.5	5.2	1236	5	Q9GTX2	Q9gtx2 plasmodium
41	206	5.1	17352	5	Q95YM2	Q95ym2 procambarus
42	204.5	5.1	3111	5	Q9VH10	Q9vh10 drosophila
43	204	5.1	976	12	Q9DUN0	Q9dun0 kaposi's sa
44	203	5.0	1233	5	Q8IJ56	Q8ij56 plasmodium
45	199	4.9	1271	5	Q25860	Q25860 plasmodium

ALIGNMENTS

RESULT 1
Q9Y4J0
ID Q9Y4J0 PRELIMINARY; PRT; 758 AA.
AC Q9Y4J0;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aspartyl(asparaginyl)beta-hydroxylase.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96420598; PubMed=8823296;
RA Lavaissiere L., Jia S., Nishiyama M., de la Monte S., Stern A.M.,
RA Wands J.R., Friedman P.A.;
RT "Overexpression of human aspartyl(asparaginyl)beta-hydroxylase in
RT hepatocellular carcinoma and cholangiocarcinoma.";
RL J. Clin. Invest. 98:1313-1323(1996).
DR EMBL; S83325; AAB50779.1; -.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR008940; Prenyl_trans.
DR InterPro; IPR001440; TPR.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
SQ SEQUENCE 758 AA; 85890 MW; 4AF6F0AB4500AF0C CRC64;

Query Match 100.0%; Score 4022; DB 4; Length 758;
Best Local Similarity 100.0%; Pred. No. 9.1e-232;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
Db        61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy       121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy       181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy       241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy       301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy       361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy       421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy       481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy       541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy       601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy       661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy       721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
             |||||||||||||||||||||||||||||||||||||
Db       721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
```

RESULT 2
Q9EPA6
ID Q9EPA6 PRELIMINARY; PRT; 739 AA.
AC Q9EPA6;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aspartyl beta-hydroxylase 6.6 kb transcript (Aspartyl beta-hydroxylase
DE 4.5 kb transcript).
GN 2310005F16RIK OR ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=BALB/c; TISSUE=Liver;
RX MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta -Hydroxylase (Asph) and an Evolutionarily Conserved
RT Isoform of Asph Missing the Catalytic Domain Share Exons with
RT Junctin.";
RL J. Biol. Chem. 275:39543-39554(2000).
DR EMBL; AF289487; AAG40809.1; -.
DR EMBL; AF289486; AAG40808.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR001440; TPR.
DR InterPro; IPR008941; TPR-like.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
SQ SEQUENCE 739 AA; 82841 MW; 4DF9F642512CA4EB CRC64;

```
  Query Match             79.8%;  Score 3209.5;  DB 11;  Length 739;
  Best Local Similarity   80.7%;  Pred. No. 2.8e-183;
  Matches  619;  Conservative   41;  Mismatches   70;  Indels   37;  Gaps    8;

Qy         1 MAQRKNAK-SSGNSSSSGSGSGS--------TSAGSSSPGARRETKHGGHKNGRKGGLSG 51
             || |||||  ||||||||||||          :|  |||||||||:|||||||||:||:||
Db         1 MAPRKNAKGGGGNSSSSGSGSGSGSPSTGSSGSSSSPGARREAKHGGHKNGRRGGISG 60

Qy        52 TSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKE 111
             ||||||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db        61 GSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGVYDADGDGDFDVDDAKVLLGLKE 120

Qy       112 RSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQ 171
             || ||   || ||| | | ||| |  |: ||:||| |||||||| | || :|    || :
Db       121 RSPSERTFPP-EAETHAELEEQAPEGADIQNVEDEVKEQIQSLLQESVHTDH----DL-E 174

Qy       172 EDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMS 231
             || |||| | ::||  || ||||| |||   ||| | :||||:| ||:   ||||| :
Db       175 ADGLAGEPQPEVEDFLTVTDSDDRFEDLEPGTVHEEIEDTYHVEDTASQNHPNDMEEMTN 234

Qy       232 EQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDS 291
             |||| | || | :   |    ::| :| |:|  |||| |:||:||:        || ::||
Db       235 EQENSDPSEAVTDAGVLLPHAEEVRHQDYDE-PVYEPSEHEGVEIS-------DNTIDDS 286

Qy       292 QVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKL 351
              :| ||:::  ||||: ||                |||||||||||||||||||||||||
Db       287 SIISEEINVASVEEQQDTPP--------------VKKKKPKLLNKFDKTIKAELDAAEKL 332

Qy       352 RKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASL 411
             |||||||||||||:|||||||||||||||||||||||||:||||||| ||||||| | |
Db       333 RKRGKIEEAVNAFEELVRKYPQSPRARYGKAQCEDDLAEKQRSNEVLRRAIETYQEAADL 392

Qy       412 PDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDND 471
             || | ||:|||||||||:|||||||||||||||||||||:||:|||||||||||:||||
Db       393 PDAPTDLVKLSLKRRSERQQFLGHMRGSLLTLQRLVQLFPSDTTLKNDLGVGYLLLGDND 452

Qy       472 NAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHL 531
             :|||||||||:|||||||||||||||||||||||||:|||||||||||||||||||||||
Db       453 SAKKVYEEVLNVTPNDGFAKVHYGFILKAQNKISESIPYLKEGIESGDPGTDDGRFYFHL 512

Qy       532 GDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLE 591
             |||||||||||||||||||||||||||||||||||||||||||||||:|||||||||||
Db       513 GDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPRETGYTELVKSLE 572

Qy       592 RNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLL 651
             |||||||||| ||||||||||||||||||||||||||||||:|||||||||| ||
Db       573 RNWKLIRDEGLMVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRKNENACKGAPKTCALL 632

Qy       652 EKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTW 711
             ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       633 EKFSETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTW 692

Qy       712 EEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
             |||||||||||||||||||||||||||||||||||||||||||||||
Db       693 EEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 739
```

RESULT 3
Q8BSY0
ID Q8BSY0 PRELIMINARY; PRT; 741 AA.
AC Q8BSY0;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

```
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Aspartate-beta-hydroxylase.
GN   ASPH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Ovary, and Uterus;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK030293; BAC26882.1; -.
DR   MGD; MGI:1914186; Asph.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   InterPro; IPR007803; Asp_Arg_Hydrox.
DR   InterPro; IPR001440; TPR.
DR   InterPro; IPR008941; TPR-like.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
DR   Pfam; PF05118; Asp_Arg_Hydrox; 1.
DR   SMART; SM00028; TPR; 2.
SQ   SEQUENCE   741 AA;  83042 MW;  0660A6A5E34418C8 CRC64;

  Query Match             79.8%;  Score 3208.5;  DB 11;  Length 741;
  Best Local Similarity   80.6%;  Pred. No. 3.2e-183;
  Matches  619;  Conservative   41;  Mismatches   71;  Indels   37;  Gaps    8;

Qy          1 MAQRKNAK-SSGNSSSSGSGSGS--------TSAGSSSPGARRETKHGGHKNGRKGGLSG 51
              || |||||   |||||||||||||        :|  |||||||| |||||||||:||:||
Db          1 MAPRKNAKGGGGNSSSSGSGSGSGSPSTGSSGSSSSPGARREAKHGGHKNGRRGGISG 60

Qy         52 TSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKE 111
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db         61 GSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGVYDADGDGDFDVDDAKVLLGLKE 120

Qy        112 RSTSEPAVPP-EEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQ 170
              || ||   || |||| | | ||| |   |: ||:||| |||||||| | || :|    ||
Db        121 RSPSERTFPPEEEAETHAELEEQAPEGADIQNVEDEVKEQIQSLLQESVHTDH----DL- 175

Qy        171 QEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMM 230
              : ||  |||| | ::|| || ||||| |||   ||| | :||||:| ||:   |||||
Db        176 EADGLAGEPQPEVEDFLTVTDSDDRFEDLEPGTVHEEIEDTYHVEDTASQNHPNDMEEMT 235

Qy        231 SEQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVED 290
              :|||| | || | :   |    ::| :| |:|  |||| |:||: |:        || ::|
Db        236 NEQENSDPSEAVTDAGVLLPHAEEVRHQDYDE-PVYEPSEHEGVAIS-------DNTIDD 287

Qy        291 SQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEK 350
              | :| ||:::  ||||: ||                |||||||||||||||||||||||
Db        288 SSIISEEINVASVEEQQDTPP----------------VKKKKPKLLNKFDKTIKAELDAAEK 333

Qy        351 LRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVAS 410
              |||||||||||||:|||||||||||||||||||||||:|||||| |||||| |
Db        334 LRKRGKIEEAVNAFEELVRKYPQSPRARYGKAQCEDDLAEKQRSNEVLRRAIETYQEAAD 393

Qy        411 LPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDN 470
              ||| | ||:||||||||:|||||||||||||||||||||:||:||||||||||:|||
Db        394 LPDAPTDLVKLSLKRRSERQQFLGHMRGSLLTLQRLVQLFPSDTTLKNDLGVGYLLLGDN 453

Qy        471 DNAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFH 530
              |:|||||||:|||||||||||||||||:|||||||||||||||||||||||||||||||
Db        454 DSAKKVYEEVLNVTPNDGFAKVHYGFILKAQNKISESIPYLKEGIESGDPGTDDGRFYFH 513

Qy        531 LGDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSL 590
              ||||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        514 LGDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPRETGYTELVKSL 573

Qy        591 ERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTL 650
              |||||||||||| ||||||||||||||||||||||||||||:|||||||||||| |
Db        574 ERNWKLIRDEGLMVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRKNENACKGAPKTCAL 633

Qy        651 LEKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRT 710
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        634 LEKFSETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRT 693

Qy        711 WEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||||||||||||||
Db        694 WEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 741


RESULT 4
Q8BQK0
ID   Q8BQK0      PRELIMINARY;      PRT;   725 AA.
AC   Q8BQK0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Aspartate-beta-hydroxylase.
GN   ASPH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Body;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573(2002).
DR   EMBL; AK049506; BAC33783.1; -.
DR   MGD; MGI:1914186; Asph.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   InterPro; IPR007803; Asp_Arg_Hydrox.
DR   InterPro; IPR001440; TPR.
DR   InterPro; IPR008941; TPR-like.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
DR   Pfam; PF05118; Asp_Arg_Hydrox; 1.
DR   SMART; SM00028; TPR; 2.
SQ   SEQUENCE   725 AA;  81492 MW;  F3CE979F8FC9C3D5 CRC64;

  Query Match             79.0%;  Score 3175.5;  DB 11;  Length 725;
  Best Local Similarity   79.8%;  Pred. No. 2.9e-181;
  Matches  613;  Conservative   41;  Mismatches   61;  Indels   53;  Gaps    9;

Qy          1 MAQRKNAK-SSGNSSSSGSGSGS--------TSAGSSSPGARRETKHGGHKNGRKGGLSG 51
              || |||||   |||||||||||||        :|  |||||||| |||||||||:||:||
Db          1 MAPRKNAKGGGGNSSSSGSGSGSGSPSTGSSGSSSSPGARREAKHGGHKNGRRGGISG 60

Qy         52 TSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKE 111
              ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||||
Db         61 GSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGVYDADGDGDFDVDDAKVLLGLKE 120

Qy        112 RSTSEPAVPP-EEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQ 170
              || ||   || |||| | | ||| |   |: ||:||| |||||||| | || :|    ||
Db        121 RSPSERTFPPEEEAETHAELEEQAPEGADIQNVEDEVKEQIQSLLQESVHTDH----DL- 175

Qy        171 QEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMM 230
              : ||  |||| | ::|| || ||||| |||   ||| | :||||:| ||:   |||||
Db        176 EADGLAGEPQPEVEDFLTVTDSDDRFEDLEPGTVHEEIEDTYHVEDTASQNHPNDMEEMT 235

Qy        231 SEQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVED 290
              :|||| | || | :   |    ::| :| |:|  |||| |:||: |:        || ::|
Db        236 NEQEN----------------SEEVRHQDYDE-PVYEPSEHEGVAIS-------DNTIDD 271

Qy        291 SQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEK 350
              | :| ||:::  ||||: ||                |||||||||||||||||||||||
Db        272 SSIISEEINVASVEEQQDTPP--------------VKKKKPKLLNKFDKTIKAELDAAEK 317

Qy        351 LRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVAS 410
```

```
                |||||||||||||||:|||||||||||||||||||||||||||:|||||||  |||||||  |
Db        318 LRKRGKIEEAVNAFEELVRKYPQSPRARYGKAQCEDDLAEKQRSNEVLRRAIETYQEAAD 377

Qy        411 LPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDN 470
              |||  |  ||:|||||||:|||||||||||||||||||||:||:|||||||| |:|||
Db        378 LPDAPTDLVKLSLKRRSERQQFLGHMRGSLLTLQRLVQLFPSDTTLKNDLGVGYHLLGDN 437

Qy        471 DNAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFH 530
              |:|||||||||:|||||||||||||||||||||:|||||||||||||||||||||||||||
Db        438 DSAKKVYEEVLNVTPNDGFAKVHYGFILKAQNKISESIPYLKEGIESGDPGTDDGRFYFH 497

Qy        531 LGDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSL 590
              ||||||||||||||||||||||||||||||||||||||||||||:||||||||||||
Db        498 LGDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPRETGYTELVKSL 557

Qy        591 ERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTL 650
              |||||||||||| ||||||||||||||||||||||||||||||:||||||||||| |
Db        558 ERNWKLIRDEGLMVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRKNENACKGAPKTCAL 617

Qy        651 LEKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRT 710
              |||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        618 LEKFSETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRT 677

Qy        711 WEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              ||||||||||||||||||||||||||||||||||||||||||||||||
Db        678 WEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 725
```

RESULT 5
Q8CBM2
ID Q8CBM2 PRELIMINARY; PRT; 658 AA.
AC Q8CBM2;
DT 01-MAR-2003 (TrEMBLrel. 23, Created)
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aspartate-beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C57BL/6J; TISSUE=Urinary bladder;
RX MEDLINE=22354683; PubMed=12466851;
RA The FANTOM Consortium,
RA the RIKEN Genome Exploration Research Group Phase I & II Team;
RT "Analysis of the mouse transcriptome based on functional annotation of
RT 60,770 full-length cDNAs.";
RL Nature 420:563-573(2002).
DR EMBL; AK035735; BAC29171.1; -.
DR MGD; MGI:1914186; Asph.
DR InterPro; IPR007943; Asp-B-hydro_N.
DR InterPro; IPR007803; Asp_Arg_Hydrox.
DR InterPro; IPR008940; Prenyl_trans.
DR InterPro; IPR001440; TPR.
DR Pfam; PF05279; Asp-B-Hydro_N; 1.
DR Pfam; PF05118; Asp_Arg_Hydrox; 1.
DR SMART; SM00028; TPR; 2.
SQ SEQUENCE 658 AA; 75127 MW; 2BA0A5C8E06801C8 CRC64;

Query Match 73.3%; Score 2950; DB 11; Length 658;
Best Local Similarity 80.7%; Pred. No. 7.3e-168;
Matches 566; Conservative 38; Mismatches 53; Indels 44; Gaps 7;

```
Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              |||||||||||||||||||||||||||||:|||||||||||||||||||||| ||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGVYDADGDGDFDVDDAKVLLGLKERSPSERT 60

Qy        119 VPP-EEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTG 177
                || |||| | | ||| |  |: ||:||| |||||||| | || :|    || : ||  |
Db         61 FPPEEEEAETHAELEEQAPEGADIQNVEDEVKEQIQSLLQESVHTDH----DL-EADGLAG 115

Qy        178 EPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPD 237
              ||| | ::||  || ||||| |||   ||| | :||||:| ||:   ||||| :||||
Db        116 EPQPEVEDFLTVTDSDDRFEDLEPGTVHEEIEDTYHVEDTASQNHPNDMEEMTNEQEN-- 173

Qy        238 SSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEE 297
                          :::| :| |:|  |||| |:||: |:       || ::|| :| ||
Db        174 -------------SEEVRHQDYDE-PVYEPSEHEGVAIS-------DNTIDDSSIISEE 211

Qy        298 VSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKI 357
              :::  ||||: ||              |||||||||||||||||||||||||||||||
Db        212 INVASVEEQQDTPP-------------VKKKKPKLLNKFDKTIKAELDAAEKLRKRGKI 257

Qy        358 EEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPAD 417
              |||||||:|||||||||||||||||||||||||||||:||||||| |||||| | ||| | |
Db        258 EEAVNAFEELVRKYPQSPRARYGKAQCEDDLAEKQRSNEVLRRAIETYQEAADLPDAPTD 317

Qy        418 LLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVY 477
              |:||||||:|||||||||||||||||||||||:||:||||||||||:||||:|||||
Db        318 LVKLSLKRRSERQQFLGHMRGSLLTLQRLVQLFPSDTTLKNDLGVGYLLLGDNDSAKKVY 377

Qy        478 EEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQR 537
              ||||:|||||||||||||||||||:|||||||||||||||||||||||||||||||||||
Db        378 EEVLNVTPNDGFAKVHYGFILKAQNKISESIPYLKEGIESGDPGTDDGRFYFHLGDAMQR 437

Qy        538 VGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLI 597
              ||||||||||||||||||||||||||||||||||||||||||:|||||||||||||||||
Db        438 VGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPRETGYTELVKSLERNWKLI 497

Qy        598 RDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPET 657
              ||||| ||||||||||||||||||||||||||||||:|||||||||||| ||||| ||
Db        498 RDEGLMVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRKNENACKGAPKTCALLEKFSET 557

Qy        658 TGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVL 717
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        558 TGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVL 617

Qy        718 IFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||||||
Db        618 IFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 658
```

RESULT 6
Q9EQ66
ID Q9EQ66 PRELIMINARY; PRT; 689 AA.
AC Q9EQ66;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE Aspartly beta-hydroxylase.
GN ASPH.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20564328; PubMed=10956665;
RA Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA Friedman P.A.;
RT "Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved
RT isoform of asph missing the catalytic domain share exons with
RT junctin.";
RL J. Biol. Chem. 275:39543-39554(2000).
RN [2]
RP SEQUENCE FROM N.A.
RA Henderson N.L., Dinchuk J.E., Burn T.C., Hollis G.F., Friedman P.A.;
RL Submitted (MAR-2001) to the EMBL/GenBank/DDBJ databases.
DR EMBL; AF289215; AAG39913.1; -.
DR EMBL; AF289205; AAG39913.1; JOINED.
DR EMBL; AF289206; AAG39913.1; JOINED.
DR EMBL; AF289207; AAG39913.1; JOINED.
DR EMBL; AF289208; AAG39913.1; JOINED.

```
DR   EMBL; AF289209; AAG39913.1; JOINED.
DR   EMBL; AF289210; AAG39913.1; JOINED.
DR   EMBL; AF289211; AAG39913.1; JOINED.
DR   EMBL; AF289212; AAG39913.1; JOINED.
DR   EMBL; AF289213; AAG39913.1; JOINED.
DR   EMBL; AF289214; AAG39913.1; JOINED.
DR   MGD; MGI:1914186; Asph.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   InterPro; IPR007803; Asp_Arg_Hydrox.
DR   InterPro; IPR001440; TPR.
DR   InterPro; IPR008941; TPR-like.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
DR   Pfam; PF05118; Asp_Arg_Hydrox; 1.
SQ   SEQUENCE   689 AA;  77319 MW;  9CB916DF109F432C CRC64;

  Query Match           72.6%;  Score 2919.5;  DB 11;  Length 689;
  Best Local Similarity   74.7%;  Pred. No. 5.2e-166;
  Matches  573;  Conservative   39;  Mismatches   68;  Indels   87;  Gaps    9;

Qy          1 MAQRKNAK-SSGNSSSSGSGSGS--------TSAGSSSPGARRETKHGGHKNGRKGGLSG 51
              || |||||   |||||||||||||        :|  ||||||||
Db          1 MAPRKNAKGGGGNSSSSGSGSGSGSGSPSTGSSGSSSSPGARR----------------- 43

Qy         52 TSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKE 111
                                          ||||:|||||||||||||||||||||
Db         44 --------------------------------GKLGVYDADGDGDFDVDDAKVLLGLKE 70

Qy        112 RSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQ 171
              || ||   || ||| | | ||| |  |: ||:||| |||||||| | || :|    || :
Db         71 RSPSERTFPP-EAETHAELEEQAPEGADIQNVEDEVKEQIQSLLQESVHTDH----DL-E 124

Qy        172 EDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMS 231
              ||  |||| | ::||  || ||||| |||   ||| | :||||:| ||:   ||||| :
Db        125 ADGLAGEPQPEVEDFLTVTDSDDRFEDLEPGTVHEEIEDTYHVEDTASQNHPNDMEEMTN 184

Qy        232 EQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDS 291
              |||| | || | :   |    ::| :| |:|  |||| |:||:||:       || ::||
Db        185 EQENSDPSEAVTDAGVLLPHAEEVRHQDYDE-PVYEPSEHEGVEIS-------DNTIDDS 236

Qy        292 QVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKL 351
              :| ||:::  ||||: ||               |||||||||||||||||||||||||
Db        237 SIISEEINVASVEEQQDTPP-------------VKKKKPKLLNKFDKTIKAELDAAEKL 282

Qy        352 RKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASL 411
              |||||||||||||:|||||||||||||||||||||||:|||||| ||||||| | |
Db        283 RKRGKIEEAVNAFEELVRKYPQSPRARYGKAQCEDDLAEKQRSNEVLRRAIETYQEAADL 342

Qy        412 PDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDND 471
              || | ||:|||||||:|||||||||||||||||||||:||:|||||||||||:||||
Db        343 PDAPTDLVKLSLKRRSERQQFLGHMRGSLLTLQRLVQLFPSDTTLKNDLGVGYLLLGDND 402

Qy        472 NAKKVYEEVLSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHL 531
              :|||||||:|||||||||||||||:|||||||||||||||||||||||||||||||||||
Db        403 SAKKVYEEVLNVTPNDGFAKVHYGFILKAQNKISESIPYLKEGIESGDPGTDDGRFYFHL 462

Qy        532 GDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLE 591
              |||||||||||||||||||||||||||||||||||||||||||||:|||||||||||||
Db        463 GDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPRETGYTELVKSLE 522

Qy        592 RNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLL 651
              |||||||||| ||||||||||||||||||||||||||||||:|||||||||||| ||
Db        523 RNWKLIRDEGLMVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRKNENACKGAPKTCALL 582

Qy        652 EKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTW 711
              ||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        583 EKFSETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTW 642

Qy        712 EEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||||||||||||
Db        643 EEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 689


RESULT 7
```

```
Q9H2C4
ID   Q9H2C4      PRELIMINARY;      PRT;   313 AA.
AC   Q9H2C4;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Aspartyl beta-hydroxylase 2.8 kb transcript.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Liver;
RX   MEDLINE=20564328; PubMed=10956665;
RA   Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA   O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA   Friedman P.A.;
RT   "Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved
RT   isoform of asph missing the catalytic domain share exons with
RT   junctin.";
RL   J. Biol. Chem. 275:39543-39554(2000).
DR   EMBL; AF289489; AAG40811.1; -.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ   SEQUENCE   313 AA;  34646 MW;  7885A18B81CD6D0D CRC64;

  Query Match           40.7%;  Score 1636;  DB 4;  Length 313;
  Best Local Similarity   99.7%;  Pred. No. 7.3e-90;
  Matches  312;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPET 313
              |||||||||||:|
Db        301 FPVEEQQEVPPDT 313


RESULT 8
Q9H291
ID   Q9H291      PRELIMINARY;      PRT;   299 AA.
AC   Q9H291;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Junctate.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20564329; PubMed=11007777;
RA   Treves S., Feriotto G., Moccagatta L., Gambari R., Zorzato F.;
RT   "Molecular Cloning, Expression, Functional Characterization,
```

```
RT   Chromosomal Localization, and Gene Structure of Junctate, a Novel
RT   Integral Calcium Binding Protein of Sarco(endo)plasmic Reticulum
RT   Membrane.";
RL   J. Biol. Chem. 275:39555-39568(2000).
DR   EMBL; AF306765; AAG42257.1; -.
DR   GO; GO:0005789; C:endoplasmic reticulum membrane; NAS.
DR   GO; GO:0005509; F:calcium ion binding; NAS.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ   SEQUENCE   299 AA;  33815 MW;  658F88C34EC2CA37 CRC64;

  Query Match             36.2%;  Score 1457.5;  DB 4;  Length 299;
  Best Local Similarity   94.2%;  Pred. No. 3.1e-79;
  Matches  278;  Conservative    2;  Mismatches    0;  Indels   15;  Gaps    1;

Qy         34 RETKHGGHKNGRKGGLSGTSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVL--------- 84
              :|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          5 KETKHGGHKNGRKGGLSGTSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLAKAKDFRYN 64

Qy         85 ------GKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEA 138
                  ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         65 LSEVLQGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEA 124

Qy        139 EPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFET 198
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        125 EPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFET 184

Qy        199 LEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDTDDVTYQ 258
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        185 LEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDTDDVTYQ 244

Qy        259 VYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPET 313
              ||||||||||||||||||||||||||||||||||||||||||||||||||||:|
Db        245 VYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPDT 299


RESULT 9
Q8TB28
ID   Q8TB28      PRELIMINARY;      PRT;   270 AA.
AC   Q8TB28;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to aspartate beta-hydroxylase.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Pancreas;
RA   Strausberg R.;
RL   Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC025236; AAH25236.1; -.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ   SEQUENCE   270 AA;  29757 MW;  8551773C7272202A CRC64;

  Query Match             34.2%;  Score 1376.5;  DB 4;  Length 270;
  Best Local Similarity   85.9%;  Pred. No. 1.8e-74;
  Matches  269;  Conservative    1;  Mismatches    0;  Indels   43;  Gaps    1;

Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEH----------------- 163

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
                                ||||||||||||||||||||||||||||||||||||||||
Db        164 -------------------------ETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 197

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        198 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 257

Qy        301 FPVEEQQEVPPET 313
              |||||||||||:|
Db        258 FPVEEQQEVPPDT 270


RESULT 10
Q9GQ82
ID   Q9GQ82      PRELIMINARY;      PRT;   785 AA.
AC   Q9GQ82;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   Aspartyl beta-hydroxylase variant 1 (CG8421-PA).
GN   ASPH OR CG8421 OR CG18658.
OS   Drosophila melanogaster (Fruit fly).
OC   Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
OC   Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC   Ephydroidea; Drosophilidae; Drosophila.
OX   NCBI_TaxID=7227;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=20564328; PubMed=10956665;
RA   Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA   O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA   Friedman P.A.;
RT   "Aspartyl beta -Hydroxylase (Asph) and an Evolutionarily Conserved
RT   Isoform of Asph Missing the Catalytic Domain Share Exons with
RT   Junctin.";
RL   J. Biol. Chem. 275:39543-39554(2000).
RN   [2]
RP   SEQUENCE FROM N.A.
RC   STRAIN=Berkeley;
RX   MEDLINE=20196006; PubMed=10731132;
RA   Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
RA   Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galle R.F.,
RA   George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
RA   Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
RA   Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D.,
RA   Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
RA   Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
RA   Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
RA   Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
RA   Borkova D., Botchan M.R., Bouck J., Brokstein P., Brottier P.,
RA   Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
RA   Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
RA   de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
RA   Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P.,
RA   Durbin K.J., Evangelista C.C., Ferraz C., Ferriera S., Fleischmann W.,
RA   Fosler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
RA   Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
RA   Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
RA   Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C.,
RA   Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
RA   Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
RA   Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
RA   Liu X., Mattei B., McIntosh T.C., McLeod M.P., McPherson D.,
RA   Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrefi A.,
RA   Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
RA   Nelson D.R., Nelson K.A., Nixon K., Nusskern D.R., Pacleb J.M.,
RA   Palazzolo M., Pittman G.S., Pan S., Pollard J., Puri V., Reese M.G.,
RA   Reinert K., Remington K., Saunders R.D.C., Scheeler F., Shen H.,
RA   Shue B.C., Siden-Kiamos I., Simpson M., Skupski M.P., Smith T.,
RA   Spier E., Spradling A.C., Stapleton M., Strong R., Sun E.,
RA   Svirskas R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,
```

```
RA   Wang Z.-Y., Wassarman D.A., Weinstock G.M., Weissenbach J.,
RA   Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
RA   Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
RA   Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
RA   Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.;
RT   "The genome sequence of Drosophila melanogaster.";
RL   Science 287:2185-2195(2000).
RN   [3]
RP   SEQUENCE FROM N.A.
RA   Celniker S.E., Adams M.D., Kronmiller B., Wan K.H., Holt R.A.,
RA   Evans C.A., Gocayne J.D., Amanatides P.G., Brandon R.C., Rogers Y.,
RA   Banzon J., An H., Baldwin D., Banzon J., Beeson K.Y., Busam D.A.,
RA   Carlson J.W., Center A., Champe M., Davenport L.B., Dietz S.M.,
RA   Dodson K., Dorsett V., Doup L.E., Doyle C., Dresnek D., Farfan D.,
RA   Ferriera S., Frise E., Galle R.F., Garg N.S., George R.A.,
RA   Gonzalez M., Houck J., Hoskins R.A., Hostin D., Howland T.J.,
RA   Ibegwam C., Jalali M., Kruse D., Li P., Mattei B., Moshrefi A.,
RA   McIntosh T.C., Moy M., Murphy B., Nelson C., Nelson K.A., Nunoo J.,
RA   Pacleb J., Paragas V., Park S., Patel S., Pfeiffer B.,
RA   Phouanenavong S., Pittman G.S., Puri V., Richards S., Scheeler F.,
RA   Stapleton M., Strong R., Svirskas R., Tector C., Tyler D.,
RA   Williams S.M., Zaveri J.S., Smith H.O., Venter J.C., Rubin G.M.;
RT   "Sequencing of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A.
RA   Misra S., Crosby M.A., Matthews B.B., Bayraktaroglu L., Campbell K.,
RA   Hradecky P., Huang Y., Kaminker J.S., Prochnik S.E., Smith C.D.,
RA   Tupy J.L., Bergman C., Berman B., Carlson J.W., Celniker S.E.,
RA   Clamp M., Drysdale R., Emmert D., Frise E., de Grey A., Harris N.,
RA   Kronmiller B., Marshall B., Millburn G., Richter J., Russo S.,
RA   Searle S.M.J., Smith E., Shu S., Smutniak F., Whitfield E.,
RA   Ashburner M., Gelbart W.M., Rubin G.M., Mungall C.J., Lewis S.E.;
RT   "Annotation of Drosophila melanogaster genome.";
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A.
RA   Adams M.D., Celniker S.E., Gibbs R.A., Rubin G.M., Venter C.J.;
RL   Submitted (MAR-2000) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE FROM N.A.
RA   FlyBase;
RL   Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF289493; AAG40806.1; -.
DR   EMBL; AE003808; AAM70947.1; -.
DR   FlyBase; FBgn0034075; Asph.
DR   GO; GO:0008237; F:metallopeptidase activity; IEA.
DR   GO; GO:0008270; F:zinc ion binding; IEA.
DR   GO; GO:0006508; P:proteolysis and peptidolysis; IEA.
DR   InterPro; IPR007803; Asp_Arg_Hydrox.
DR   InterPro; IPR006025; Pept_M_Zn_BS.
DR   InterPro; IPR008940; Prenyl_trans.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF05118; Asp_Arg_Hydrox; 1.
DR   PROSITE; PS00142; ZINC_PROTEASE; 1.
SQ   SEQUENCE   785 AA;  89843 MW;  30A8DFCD6836F7F1 CRC64;

  Query Match           26.5%;  Score 1066.5;  DB 5;  Length 785;
  Best Local Similarity  33.6%;  Pred. No. 2.5e-55;
  Matches  275;  Conservative  129;  Mismatches  310;  Indels  105;  Gaps   23;

Qy          3 QRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMVIA 62
              :||: :   :   |  |   |  |:          |  | :|  |        |   | :
Db          9 KRKDKRRKRDDDESSHGVHITKMGNEDLHL-----HVHHDHGTGGHWCAKIIFFALMAV- 62

Qy         63 LLGV---------------------WTSVAVVWFDL-------VDYEEVLGK-LGIYDA 92
              |||:                     :: |   | |         | :|  |: |  :|
Db         63 LLGLVGLIIMENRGLEDLDTPLSESRFSKVFDGWVDEHRDEHDGHDVQEPSGEALDDHDE 122

Qy         93 -DGDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQI 151
               |   | : :| : |   |    |   | || ||  : | :     | :: |: | | |
Db        123 HDDHDDHEDEDEEPLTEELEEELEEEEEPTEEDEPAADEEYE-----EDEDEENNAGENI 177

Qy        152 QSLLHEMVHAEHVEGEDLQQEDG--------------PTGEPQ-QEDDEFLMATDVDDR 195
              :          ||  | |:   ::|             ||| : :|||   | | ||
Db        178 TA-----EDAEEEEEEEDNDDEGTVEATVEATTEATTEATGEYEAEEDDEDEAAAD-DDA 231

Qy        196 FETLEPEVS--------HEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDER 247
              |: |  :|         :| |    :||:|  : :|   :    ::| | :   :|::
Db        232 VESTEAPLSKAEEQEDDEDEDEEEQEIEESVEPERSQYAAQSAPAKDNNDDDD---DDDQ 288

Qy        248 LHHDTDDVTYQVYEEQAVYEPLENEGIEIITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQ 307
              :| ||     :| |  :  ||:          |||          |   ::   |: |
Db        289 DKNDADDGDDDEFESLA--QQFEND----------PEDT--------VPAKAVESEEQDQ 328

Qy        308 EVPPETNRKTDDPE---QKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEAVNAF 364
              :   |  | :       : | |:  |    | |   : ||  | :   |    :|: :|
Db        329 DQADEEEPKEEGSSWLASQIKPKEAAPAPAEKED-PFEQELRKANEEMIRENYAQALRSF 387

Qy        365 KELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLKLSLK 424
              |   :  | |  |:|:  : ||:| |||: |  ||: |:   :  :: |   :
Db        388 NTLTTNFAHEPSAHLGRARLLELLAKKERSNQRLWEAIDAYKRYLAFGELVASNQEFQTA 447

Qy        425 RRS--DRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEVLS 482
              |  :  :|||| | :   : |:   | |  |:| | : ||:: :   :||  | |
Db        448 GESCIENLRFLGHHRQATTIHELLINRLPEDPRLRNQLSLTYLMVNNLQQVEKVAVETLK 507

Qy        483 VTPNDGFAKVHYGFILKA-QNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGNK 541
              : ||:  |::|||  |:      |:::|||  :|||: || :  ||  ||: |||:  |
Db        508 LWPNNAVAQLHYGLALRQFHADYAKALPYLKYAVESGEEGTQEAFFYLSLGETMQRLSMK 567

Qy        542 -EAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
               || : |  | :| |||::|||||   |:|||:| ||||||   :: |  ||: ||||
Db        568 SEALEVYGKGVAKGFFASLYQRSLYNEPRLRAQPFWQPKETGYERQLEKLTLNWRAIRDE 627

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              |||:: :: | |  | | ||:|| | |: |: |||| :: |: || ||:|||:|||: ||
Db        628 GLALLGRS-GFFEDEAELLRDKGVWQQYELYAQGRRVKDNCRRAPITCSLLEEFPESAGC 686

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPK-EGCKIRCANETRTWEEGKVLIF 719
              ||||:|:|:|   |||||| |||||||| || |  |: |   :| | : ||| ||:: ||
Db        687 RRGQVKFSVMQAKTHVWPHCGPTNCRLRAHLTLAAPEPEKASLRVAEQERTWREGELFIF 746

Qy        720 DDSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||| : |  ||: |:|:|||:|:  |||||  |
Db        747 DDSFEHEVWHNGSQSRLVLILDMWHPQLSAAQRRSLSPI 785


RESULT 11
Q9EQ65
ID   Q9EQ65      PRELIMINARY;      PRT;   308 AA.
AC   Q9EQ65;
DT   01-MAR-2001 (TrEMBLrel. 16, Created)
DT   01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Aspartyl beta-hydroxylase 2.8 kb transcript.
GN   ASPH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c; TISSUE=Liver;
RX   MEDLINE=20564328; PubMed=10956665;
RA   Dinchuk J.E., Henderson N.L., Burn T.C., Huber R., Ho S.P., Link J.,
RA   O'Neil K.T., Focht R.J., Scully M.S., Hollis J.M., Hollis G.F.,
RA   Friedman P.A.;
RT   "Aspartyl beta -hydroxylase (Asph) and an evolutionarily conserved
RT   isoform of asph missing the catalytic domain share exons with
RT   junctin.";
RL   J. Biol. Chem. 275:39543-39554(2000).
DR   EMBL; AF289488; AAG40810.1; -.
DR   MGD; MGI:1914186; Asph.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
```

```
SQ   SEQUENCE   308 AA;  33142 MW;  F208C2C31C595282 CRC64;

  Query Match             24.8%;  Score 998.5;  DB 11;  Length 308;
  Best Local Similarity   64.3%;  Pred. No. 8e-52;
  Matches  207;  Conservative   30;  Mismatches   62;  Indels   23;  Gaps    7;

Qy          1 MAQRKNAK-SSGNSSSSGSGSGS--------TSAGSSSPGARRETKHGGHKNGRKGGLSG 51
              || |||||    ||||||||||||        :|  |||||||| ||||||||:||:||
Db          1 MAPRKNAKGGGGNSSSSGSGSGSGSGSPSTGSSGSSSSPGARREAKHGGHKNGRRGGISG 60

Qy         52 TSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKE 111
              |||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db         61 GSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGVYDADGDGDFDVDDAKVLLGLKE 120

Qy        112 RSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQ 171
              || ||   || ||| | | ||| |  |: ||:||| |||||||| | || :|    || :
Db        121 RSPSERTFPP-EAETHAELEEQAPEGADIQNVEDEVKEQIQSLLQESVHTDH----DL-E 174

Qy        172 EDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMS 231
              ||  |||| | ::||  || ||||| |||   ||| | :||||:| ||:   ||||| :
Db        175 ADGLAGEPQPEVEDFLTVTDSDDRFEDLEPGTVHEEIEDTYHVEDTASQNHPNDMEEMTN 234

Qy        232 EQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDS 291
              |||| | || | :   |    ::| :| |:|  |||| |:||:||:       || ::||
Db        235 EQENSDPSEAVTDAGVLLPHAEEVRHQDYDE-PVYEPSEHEGVEIS-------DNTIDDS 286

Qy        292 QVIVEEVSIFPVEEQQEVPPET 313
              :| ||:::  |||||: ||:|
Db        287 SIISEEINVASVEEQQDTPPDT 308


RESULT 12
Q91WG6
ID   Q91WG6      PRELIMINARY;      PRT;   292 AA.
AC   Q91WG6;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Similar to aspartate-beta-hydroxylase.
GN   ASPH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Kidney;
RA   Strausberg R.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC015281; AAH15281.1; -.
DR   MGD; MGI:1914186; Asph.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ   SEQUENCE   292 AA;  31568 MW;  B3A470A46EC24B6D CRC64;

  Query Match             24.2%;  Score 972.5;  DB 11;  Length 292;
  Best Local Similarity   62.7%;  Pred. No. 2.7e-50;
  Matches  202;  Conservative   30;  Mismatches   51;  Indels   39;  Gaps    8;

Qy          1 MAQRKNAK-SSGNSSSSGSGSGS--------TSAGSSSPGARRETKHGGHKNGRKGGLSG 51
              || |||||    ||||||||||||        :|  |||||||| ||||||||:||:||
Db          1 MAPRKNAKGGGGNSSSSGSGSGSGSGSPSTGSSGSSSSPGARREAKHGGHKNGRRGGISG 60

Qy         52 TSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKE 111
              |||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db         61 GSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGVYDADGDGDFDVDDAKVLLGLKE 120

Qy        112 RSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQ 171
              || ||   || ||| | | ||| |  |: ||:||| |||||||| | || :|    || :
Db        121 RSPSERTFPP-EAETHAELEEQAPEGADIQNVEDEVKEQIQSLLQESVHTDH----DL-E 174

Qy        172 EDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMS 231
              ||  |||| | ::||  || ||||| |||   ||| | :||||:| ||:   ||||| :
Db        175 ADGLAGEPQPEVEDFLTVTDSDDRFEDLEPGTVHEEIEDTYHVEDTASQNHPNDMEEMTN 234

Qy        232 EQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDS 291
              |||| | || | :   |    ::| :| |:|  |||| |:||:||:       || ::||
Db        235 EQENSDPSEAVTDAGVLLPHAEEVRHQDYDE-PVYEPSEHEGVEIS-------DNTIDDS 286

Qy        292 QVIVEEVSIFPVEEQQEVPPET 313
              :| ||:::  |||||: ||:|
Db        287 SIISEEINVASVEEQQDTPPDT 308


RESULT 13
Q920F9
ID   Q920F9      PRELIMINARY;      PRT;   270 AA.
AC   Q920F9;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cardiac junctate 1.
GN   ASPH.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RA   Hong C.-S., Kim D.H.;
RT   "Cloning of mouse junctin homologs.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF302653; AAL09319.2; -.
DR   PIR; JC7792; JC7792.
DR   MGD; MGI:1914186; Asph.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ   SEQUENCE   270 AA;  29949 MW;  70BB13DC01B51701 CRC64;

  Query Match             22.1%;  Score 890;  DB 11;  Length 270;
  Best Local Similarity   63.9%;  Pred. No. 2e-45;
  Matches  179;  Conservative   30;  Mismatches   57;  Indels   14;  Gaps    5;

Qy         34 RETKHGGHKNGRKGGLSGTSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDAD 93
              :| |||||||||:||:|| ||||||||||||||||||||||||||||||||:||||
Db          5 KEAKHGGHKNGRRGGISGGSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGVYDAD 64

Qy         94 GDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQS 153
              |||||||||||||||||| ||   || ||| | | ||| |  |: ||:||| ||||||
Db         65 GDGDFDVDDAKVLLGLKERSPSERTFPP-EAETHAELEEQAPEGADIQNVEDEVKEQIQS 123

Qy        154 LLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYH 213
              || | || :|    || : ||  |||| | ::||  || ||||| |||   ||| | :||
Db        124 LLQESVHTDH----DL-EADGLAGEPQPEVEDFLTVTDSDDRFEDLEPGTVHEEIEDTYH 178

Qy        214 VEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEG 273
              ||:| ||:   ||||| :|||| | || | :   |    ::| :| |:|  |||| |:||
Db        179 VEDTASQNHPNDMEEMTNEQENSDPSEAVTDAGVLLPHAEEVRHQDYDE-PVYEPSEHEG 237

Qy        274 IEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPET 313
              :||:       || ::|| :| ||:::  |||||: ||:|
Db        238 VEIS-------DNTIDDSSIISEEINVASVEEQQDTPPDT 270


RESULT 14
Q93178
ID   Q93178      PRELIMINARY;      PRT;   872 AA.
AC   Q93178; Q93876;
DT   01-FEB-1997 (TrEMBLrel. 02, Created)
DT   01-NOV-1998 (TrEMBLrel. 08, Last sequence update)
DT   01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
DE   K09A9.6 protein.
GN   K09A9.6.
OS   Caenorhabditis elegans.
OC   Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
```

```
OC   Rhabditidae; Peloderinae; Caenorhabditis.
OX   NCBI_TaxID=6239;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Swinburne J.;
RL   Submitted (AUG-1996) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=94150718; PubMed=7906398;
RA   Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,
RA   Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,
RA   Craxton M., Dear S., Du Z., Durbin R., Favello A., Fulton L.,
RA   Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA   Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA   Lightning J., Lloyd C., Mcmurray A., Mortimore B., O'Callaghan M.,
RA   Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,
RA   Smaldon N., Smith A., Sonnhammer E., Staden R., Sulston J.,
RA   Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Waterston R.,
RA   Watson A., Weinstock L., Wilkinson-Sproat J., Wohldman P.;
RT   "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT   elegans.";
RL   Nature 368:32-38(1994).
DR   EMBL; Z79601; CAB01887.1; -.
DR   EMBL; Z79596; CAB01887.1; JOINED.
DR   EMBL; Z79596; CAB01859.1; -.
DR   EMBL; Z79601; CAB01859.1; JOINED.
DR   PIR; T18861; T18861.
DR   WormPep; K09A9.6; CE11982.
DR   InterPro; IPR007803; Asp_Arg_Hydrox.
DR   InterPro; IPR008940; Prenyl_trans.
DR   InterPro; IPR001440; TPR.
DR   Pfam; PF05118; Asp_Arg_Hydrox; 1.
SQ   SEQUENCE   872 AA;  99456 MW;  8CA3E03489A9848B CRC64;

  Query Match             19.8%;  Score 798;  DB 5;  Length 872;
  Best Local Similarity   26.8%;  Pred. No. 3.1e-39;
  Matches  191;  Conservative  144;  Mismatches  278;  Indels  100;  Gaps   15;

Qy         89 IYDADGDGDFDVDDAKVLLGLKERSTSEPAVPPEEAEPHTEPEEQ-------VPVEAEPQ 141
Db        205 VEDDDDDDDEDDDDVE---------APAPQEPSRKQKAHNQREEKKDKNKKYQPVKEEPD 255

Qy        142 NIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDG------PTGEPQQEDDEFLMATDVDDR 195
Db        256 LDDDDDDEDDDDDDGDDDDDEKDDGNDAEKDDGDDDDDDDGNDDEEDGKNKKTSVEAK 315

Qy        196 FETL---------EPEVSHEETE--HSYHVEET--------------VSQDC-NQDMEE- 228
Db        316 SDKVVEEKNDGDDEPHVSHREAQRLRRQHLRESRRDNRPRQGGNRECIHQDCPNRESLKP 375

Qy        229 ----MMSEQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPE 284
Db        376 RKSLLVTKKKTKSMVERILDDEFDDEDDDD--------------EDDDSEVNK-----N 415

Qy        285 DNPVEDSQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAE 344
Db        416 DREDDDEDDVDERIS--------------------DRDSSSSYKRHAITTKEEIGFRDI 454

Qy        345 LDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIET 404
Db        455 LDRADNLVEKHQYEEAMELFDHVIAVYPASTRAYFGKARAYDIRGEIEADETDRDKAIEI 514

Qy        405 YQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGY 464
Db        515 YEKILQNSGVPDALFRQAAQRLIEKTRFRGQLHKTLTAHRYFIDRFPEELNLQTDFAISF 574

Qy        465 LLIGDNDNAKKVYEEVLSVTPNDGFAKVHYGFILKA-QNKIAESIPYLKEGIESGDPGTD 523
Db        575 VMMKRYEDARTVLKNVLANDPNHVIALAYYGYILKAHDDKVEQGVALMRKSLKNADNEIT 634

Qy        524 DGRFYFHLGDAMQRVGNK-EAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETG 582
Db        635 DPKFYYQLGHGLTTLGRKSEADAVYQKAAQMGVFMTAQQRSLYNIEGLTGRAWWAMDQTP 694

Qy        583 YTELVKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACK 642
Db        695 YSKFLKTVERQWATIRQEGMEVLKDCSDCWLDHNQQLVIDGQWKFFPIMSEQNFIKSSCE 754

Qy        643 GAPKTCTLLEKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKI 702
Db        755 RMPQTCLILQEFAASSNASKSDMHLSVLSSGASILPHCGPTNYHLQAHLGLVSPSE-ARI 813

Qy        703 RCANETRTWEEGKVLIFDDSFEHEVWQD---ASSFRLIFIVDVWHPELTPQQR 752
Db        814 RVGNETKGWRSGKFIIYDDSFEHELQFDGASSSSFRLVLTIQLWHPEVQPHQR 866


RESULT 15
Q920F8
ID   Q920F8      PRELIMINARY;      PRT;   259 AA.
AC   Q920F8;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE   Cardiac junctate 2.
GN   ASPH OR 3110001L23RIK.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RA   Hong C.-S., Kim D.H.;
RT   "Cloning of mouse junctin homologs.";
RL   Submitted (SEP-2000) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF302654; AAL09320.1; -.
DR   MGD; MGI:1914186; Asph.
DR   InterPro; IPR007943; Asp-B-hydro_N.
DR   Pfam; PF05279; Asp-B-Hydro_N; 1.
SQ   SEQUENCE   259 AA;  28455 MW;  A6740C6CF199E093 CRC64;

  Query Match             19.4%;  Score 779.5;  DB 11;  Length 259;
  Best Local Similarity   55.4%;  Pred. No. 7.5e-39;
  Matches  163;  Conservative   27;  Mismatches   51;  Indels   53;  Gaps    6;

Qy         34 RETKHGGHKNGRKGGLSGTSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDAD 93
Db          5 KEAKHGGHKNGRRGGISGGSFFTWFMVIALLGVWTSVAVVWFDLVDYEEVLGKLGVYDAD 64

Qy         94 GDGDFDVDDAKVLL--------------GLKERSTSEPAVPPEEAEPHTEPEEQVPVEAE 139
Db         65 GDGDFDVDDAKVLLEGPGGLAKRKTKAKGLKERSPSERTFPP-EAETHAELEEQAP---- 119

Qy        140 PQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQEDDEFLMATDVDDRFETL 199
Db        120 ------------------------EGADL-EADGLAGEPQPEVEDFLTVTDSDDRFEDL 153

Qy        200 EPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDTDDVTYQV 259
Db        154 EPGTVHEEIEDTYHVEDTASQNHPNDMEEMTNEQENSDPSEAVTDAGVLLPHAEEVRHQD 213

Qy        260 YEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPET 313
Db        214 YDE-PVYEPSEHEGVEIS-------DNTIDDSSIISEEINVASVEEQQDTPPDT 259


Search completed: May  5, 2004, 11:03:56
Job time : 55 secs
```

GenCore version 5.1.6
Copyright (c) 1993 - 2004 Compugen Ltd.

OM protein - protein search, using sw model

Run on: May 5, 2004, 11:01:09 ; Search time 23 Seconds
(without alignments)
1701.412 Million cell updates/sec

Title: US-09-903-216-2
Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSSGSGS.....IVDVWHPELTPQQRRSLPAI 758

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	1334	33.2	255	3	US-09-040-485-2	Sequence 2, Appli
2	1320	32.8	255	4	US-09-702-705-1806	Sequence 1806, Ap
3	1320	32.8	255	4	US-09-736-457-1806	Sequence 1806, Ap
4	1320	32.8	255	4	US-09-671-325-1806	Sequence 1806, Ap
5	199	4.9	783	6	5231168-2	Patent No. 5231168
6	197.5	4.9	320	4	US-09-252-991A-29355	Sequence 29355, A
7	197.5	4.9	1018	1	US-08-072-610-2	Sequence 2, Appli
8	197.5	4.9	1018	2	US-08-719-822B-2	Sequence 2, Appli
9	197.5	4.9	1018	3	US-09-092-458-2	Sequence 2, Appli
10	190.5	4.7	1162	2	US-08-728-323A-2	Sequence 2, Appli
11	190.5	4.7	1162	4	US-09-298-568-2	Sequence 2, Appli
12	190.5	4.7	1162	4	US-09-410-399-2	Sequence 2, Appli
13	189.5	4.7	312	4	US-09-328-352-8015	Sequence 8015, Ap
14	177.5	4.4	565	3	US-08-961-083-218	Sequence 218, App
15	177.5	4.4	565	4	US-09-536-784-218	Sequence 218, App
16	177	4.4	327	4	US-09-489-039A-10158	Sequence 10158, A
17	172.5	4.3	700	4	US-09-107-532A-5094	Sequence 5094, Ap
18	163	4.1	411	2	US-08-741-134-6	Sequence 6, Appli
19	163	4.1	2662	4	US-09-595-684B-31	Sequence 31, Appl
20	162.5	4.0	310	4	US-09-252-991A-26511	Sequence 26511, A
21	162	4.0	1282	4	US-09-543-681A-5419	Sequence 5419, Ap
22	160	4.0	688	3	US-09-141-047-8	Sequence 8, Appli
23	156.5	3.9	506	2	US-08-820-170A-19	Sequence 19, Appl
24	156.5	3.9	506	3	US-09-055-699-19	Sequence 19, Appl
25	156.5	3.9	506	3	US-09-273-565-19	Sequence 19, Appl
26	156.5	3.9	506	4	US-09-565-538-19	Sequence 19, Appl
27	156.5	3.9	506	4	US-09-661-468-19	Sequence 19, Appl
28	156.5	3.9	506	4	US-09-976-165-19	Sequence 19, Appl
29	154.5	3.8	1196	4	US-09-107-532A-3944	Sequence 3944, Ap
30	153.5	3.8	1898	1	US-08-056-200-94	Sequence 94, Appl
31	153.5	3.8	1898	2	US-08-800-644-94	Sequence 94, Appl
32	152.5	3.8	984	1	US-08-242-932-2	Sequence 2, Appli
33	152.5	3.8	984	1	US-08-714-481-2	Sequence 2, Appli
34	152.5	3.8	984	5	PCT-US95-06111-2	Sequence 2, Appli
35	151.5	3.8	1972	4	US-08-875-435B-4	Sequence 4, Appli
36	151	3.8	404	4	US-09-554-080A-3	Sequence 3, Appli
37	151	3.8	436	4	US-09-554-080A-2	Sequence 2, Appli
38	150.5	3.7	3878	4	US-09-914-259-11	Sequence 11, Appl
39	150	3.7	571	4	US-09-216-393B-327	Sequence 327, App
40	149.5	3.7	1164	3	US-08-923-992A-10	Sequence 10, Appl
41	149	3.7	967	4	US-09-914-259-21	Sequence 21, Appl
42	148.5	3.7	816	2	US-08-533-306A-6	Sequence 6, Appli
43	148.5	3.7	816	2	US-08-742-923A-6	Sequence 6, Appli
44	148	3.7	258	3	US-08-961-083-90	Sequence 90, Appl
45	148	3.7	258	4	US-09-536-784-90	Sequence 90, Appl

ALIGNMENTS

```
RESULT 1
US-09-040-485-2
; Sequence 2, Application US/09040485
; Patent No. 6166176
;  GENERAL INFORMATION:
;    APPLICANT:  Radosevich, James A.
;    TITLE OF INVENTION:  A GENE ENCODING A NOVEL MARKER FOR
;    TITLE OF INVENTION:  CANCER
;    NUMBER OF SEQUENCES:  9
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  BRINKS, HOFER, GILSON & LIONE
;      STREET:  NBC Tower - Suite 3600, 455 N. Cityfront
;      STREET:  Plaza Drive
;      CITY:  Chicago
;      STATE:  Illinois
;      COUNTRY:  USA
;      ZIP:  60611-5599
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.30
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/040,485
;      FILING DATE:  17-MAR-1998
;      CLASSIFICATION:  530
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Martin, Alice O.
;      REGISTRATION NUMBER:  35,601
;      REFERENCE/DOCKET NUMBER:  8998/3
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  312-321-4200
;      TELEFAX:  312-321-4299
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  255 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-09-040-485-2

  Query Match             33.2%;  Score 1334;  DB 3;  Length 255;
  Best Local Similarity   99.6%;  Pred. No. 1.6e-100;
  Matches  254;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
```

```
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy         179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy         239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 240

Qy         299 SIFPVEEQQEVPPET 313
                |||||||||||||:|
Db         241 SIFPVEEQQEVPPDT 255


RESULT 2
US-09-702-705-1806
; Sequence 1806, Application US/09702705
; Patent No. 6504010
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C14
;  CURRENT APPLICATION NUMBER: US/09/702,705
;  CURRENT FILING DATE:  2000-10-30
;  NUMBER OF SEQ ID NOS:  1833
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1806
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-702-705-1806

  Query Match             32.8%;  Score 1320;  DB 4;  Length 255;
  Best Local Similarity   98.8%;  Pred. No. 2.1e-99;
  Matches  252;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy         119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy         179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
                ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db         121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy         239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
                |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy         299 SIFPVEEQQEVPPET 313
                |||||||||||||:|
Db         241 SIFPVEEQQEVPPDT 255


RESULT 3
US-09-736-457-1806
; Sequence 1806, Application US/09736457
; Patent No. 6509448
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1806
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-736-457-1806

  Query Match             32.8%;  Score 1320;  DB 4;  Length 255;
  Best Local Similarity   98.8%;  Pred. No. 2.1e-99;
  Matches  252;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy          59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy         119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
                ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy         179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
                ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db         121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy         239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
                |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db         181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy         299 SIFPVEEQQEVPPET 313
                |||||||||||||:|
Db         241 SIFPVEEQQEVPPDT 255


RESULT 4
US-09-671-325-1806
; Sequence 1806, Application US/09671325
; Patent No. 6667154
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C12
;  CURRENT APPLICATION NUMBER: US/09/671,325
;  CURRENT FILING DATE:  2000-09-26
;  NUMBER OF SEQ ID NOS:  1825
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1806
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
```

US-09-671-325-1806

Query Match 32.8%; Score 1320; DB 4; Length 255;
Best Local Similarity 98.8%; Pred. No. 2.1e-99;
Matches 252; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

```
Qy          59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db           1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy         119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy         179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy         239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
               |||||||||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db         181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy         299 SIFPVEEQQEVPPET 313
               |||||||||||||:|
Db         241 SIFPVEEQQEVPPDT 255
```

RESULT 5
5231168-2
;Patent No. 5231168
; APPLICANT: DZIEGIEL, MORTEN;BORRE, MARTIN;JEPSEN, SOREN;
;VUUST, JENS;RIENECK, KLAUS;WIND, ANNETTE;JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
;SEQ ID NO:2:
; LENGTH: 783
5231168-2

Query Match 4.9%; Score 199; DB 6; Length 783;
Best Local Similarity 20.6%; Pred. No. 1.6e-07;
Matches 143; Conservative 134; Mismatches 237; Indels 180; Gaps 36;

```
Qy          81 EEVLGKLGIYDADGDGDF----------DVDDAKVL---LGLKERSTSEPAVPPE--EAE 125
               :|  |: |: | : :|||          | :|:::    |   |:| ||||   |   :
Db         119 DEESGESGLVDNE-EGDFEEPNHEEFEPDQNDSELSENELVESEKSVSEPAEHVEIVSEK 177

Qy         126 PHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTG-------- 177
                 :|| | | : :|     | | ::|:      :|    |  :::|||
Db         178 SVSEPAEHVEIVSEKST--SEPAEHVESV------SEQSNNEPSEKKDGPVPSKPFEEIE 229

Qy         178 ----EPQQEDDEFLMATDVDDRFETLEP---------EVSHEETEHSYHVEETVSQDCNQ 224
                   :|:  | : :    || :    ||          :|  || :|:  |:  | :  :
Db         230 KVDVQPKIVDLQIIEPNFVDSQPNPQEPVEPSFVKIEKVPSEENKHA-SVDPEVKE--KE 286

Qy         225 DMEEMMSEQENPDSSE---PVVEDERLHHDTDDV-TYQVYEEQAVYEPLENEGIEITEVT 280
               :: |:: |::|    |    || |||       :|| | |:  :   :|   | :|: |:
Db         287 NVSEVVEEKQNSQESVEEIPVNEDE-----FEDVHTEQLDLDHKTVDP---EIVEVEEIP 338

Qy         281 APPEDNPVEDSQVI-VEEVSIFP----------------------VEEQQEVPPETNRK 316
               :   :| |   ::: :|||  ||                     : | :|: || ::
Db         339 SELHENEVAHPEIVEIEEV--FPEPNQNNEFQEINEDDKSAHIQHEIVEVEEILPEDDKN 396

Qy         317 TDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAE------KLRKRGKIEEAVNAFKELVRK 370
                    :  :|::  |:      ||  | : :  |      :  |  |:|  :   :|:: :
Db         397 EKVEHEIVEVEEILPE-----DKNEKGQHEIVEVEEILPEDDKNEKVEHEIVEVEEILPE 451

Qy         371 YPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPAD--------LLKLS 422
               :: : ::   : |: | | :  || :   |  :|:     :| |         ::::
Db         452 -DKNEKGQHEIVEVEEILPEDK--NEKVEHEIVEVEEI-----LPEDKNEKGQHEIVEVE 503

Qy         423 LKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLG------VGYLLIGDNDNAKKV 476
                  |: : : |        :  : :: | |   ||: |       |  :|  :: | |
Db         504 EILPEDKNEKVQH------EIVEVEEILPED---KNEKGQHEIVEVEEILPEEDKNEKGQ 554

Qy         477 YE--EVLSVTPNDGFAKVHYGFI-----------LKAQNKIAESIPYLKEGIESGDPGTD 523
               :|  ||  : | |   || :  :             | |::| |    | | :|
Db         555 HEIVEVEEILPEDKNEKVQHEIVEVEEILPEDKNEKVQHEIVEVEEILPEIVE------- 607

Qy         524 DGRFYFHLGDAMQRVGNKEAYKWYELGHKRGHFASVWQRSLYN---VNGLKAQPWWTPK- 579
                     : :   :  | |  :  :   |:  | || ::::     |  |      |||
Db         608 -------IEEVPSQTNNNENIETIKPEEKKNEF-SVEEKAIPQEPVVPTLNENENVTPKP 659

Qy         580 ---ETGYTELVKSLERNWKLIRD-----EGLAVMDKAKGL--FLPEDENLREKGDWSQFT 629
                  |:   ::|:      |::::     :   |:|  | :   : ||:|  |  |   |
Db         660 SEGESTKPDIVQI-----KIVQENKPNKKETPVVDGPKHVEQNIQEDDNDEEDDDDIDFE 714

Qy         630 LW-----QQGRRNENACKGAPKTCTLLEKFPETT 658
                     ::   |:|  | :  |    :|| : :
Db         715 GLSRKDDEKDSSNKNKKKSSFITYISTKKFKKVS 748
```

RESULT 6
US-09-252-991A-29355
; Sequence 29355, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 107196.136
; CURRENT APPLICATION NUMBER: US/09/252,991A
; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 29355
; LENGTH: 320
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-29355

Query Match 4.9%; Score 197.5; DB 4; Length 320;
Best Local Similarity 25.9%; Pred. No. 5.7e-08;
Matches 58; Conservative 34; Mismatches 85; Indels 47; Gaps 9;

```
Qy         551 HKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTEL--------------------VKSL 590
               | ||      : | | | :      |: |    | |                      :  |
Db          31 HFRGRARLPFLRQLVNHSA-----WFAP----YNSLMYLFSSVPSKPYLDRSRFPELDEL 81

Qy         591 ERNWKLIRDEGLAVMDKAKGLFLPEDENLREKG-------DWSQFTL-WQQGRRNENACK 642
               : ||: ||:| | : |:   ::    |  | |       | :| | | |  |   :| :
Db          82 KNNWQTIREEALNLFDEG---YIRAALNNNEAGFGSFFKKGWKRFYLTWYDGPL-PSAQQ 137

Qy         643 GAPKTCTLLEKFPETTGCRRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIP-KEGCK 701
                 |||  |: : |   |        ::::  |:|: ||  |    || ||||  |  : |:
Db         138 LCPKTVELVSRIPNVKGA-----MFTLLPGGSHLNPHRDPFGGSLRYHLGLSTPNSDNCR 192

Qy         702 IRCANETRTWEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVWHP 745
               |    :   | :|: ::||::| | |  :    |:|   |:  |
Db         193 IYVDGQPYAWRDGEDVMFDETFVHWVKNETEQTRVILFCDIERP 236
```

RESULT 7
US-08-072-610-2
; Sequence 2, Application US/08072610
; Patent No. 5532133
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 2

```
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby and Darby
;      STREET:  805 Third Ave.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022-7513
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/072,610
;      FILING DATE:  19930602
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  5986/07686
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)527-7700
;      TELEFAX:  (212)753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1018 amino acids
;      TYPE:  AMINO ACID
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  C-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium vivax
;    IMMEDIATE SOURCE:
;      CLONE:  PvMB3.3.1
US-08-072-610-2

  Query Match             4.9%;  Score 197.5;  DB 1;  Length 1018;
  Best Local Similarity   23.2%;  Pred. No. 3.1e-07;
  Matches  128;  Conservative   86;  Mismatches  205;  Indels  133;  Gaps   27;

Qy        109 LKERSTSEPAVPPEEAEP-----HTEPEEQVPVEAEPQNIED--------EAKEQIQSLL 155
              ||:    |    | :  |         | ::|| |    |  ||          || |: : :|
Db        533 LKDPDAGEAVTVPSKEAPVQVPVAVGPAQEVPTEELMQLQEDDFELEGTAEAPEEGELVL 592

Qy        156 HEMVHAEHVEGEDLQQEDGP-TGEPQQED--DEFLMATDVDDRFETLEPEVSHEETEHSY 212
                  |||     |: |  ||| : :  :| | ||  || ||  ||  : || | :
Db        593 ---------EGEGEPTEEEPREGEPTEGEVPEEELEATPEDD-FELEEP--TGEEVEETV 640

Qy        213 HVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENE 272
                |||       :::||: :| |                  : ::|  :| | : | | :|
Db        641 EGEETAE---GEEVEEVPAEVE---------------EVEEVPAEVEEVEEVPEEVEEV 681

Qy        273 GIEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPK 332
              |: ||   ||:          ||||     || :||| |     :: |:  :|:: :
Db        682 PAEVEEVEEVPEE---------VEEVP----EEVEEVPEEVEEVPEEVEEVEEVEEVEVP 728

Qy        333 LLNKFDKTIKAELDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKR 392
              : : :      | :  |:: :  : || |      |      |      :  | : :|:|
Db        729 AVVEVEVPAVVEEEVPEEVEEEEEEEEPVEEEDVLQLVIPSEEDIQLDKPK-KDELGSGI 787

Qy        393 RS------NEVLRGAIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRL 446
               |       :| :  :|  :| |  |  | |       | |   ::|  ::|      :||
Db        788 LSIIDMHYQDVPKEFMEEEEETAVYPLKPEDF----AKEDSQSTEWLTFIQGLEGDWERL 843

Qy        447 VQLFPNDTSL----------KNDLGVGYLLIGDNDNAKKVYEEVLSVTPND--GFAKVHY 494
                : ||          :|   |:| : :|  ::  | :: |     |  |  |  :
Db        844 ------EVSLNKARERWMEQRNKEWAGWLRLIENKWSE--YSQI-STKGKDPAGLRKREW 894

Qy        495 G-------FILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGNKEAYKWY 547
                    |  : :::|     :||: :      |     |   : | | :  ||:  :|
Db        895 SDEKWKKWFKAEVKSQIDS---HLKKWMND----THSNLFKILVKD-MSQFENKKTKEWL 946

Qy        548 -------ELGHKRGHF------------ASVWQRSLYNVNGLKAQ--PWWTPKETGY-- 583
                      | |:   |            :  | |:  |:|  : :   |:  ||  |
Db        947 MNHWKKNERGYGSESPEVMTTSKLLNVAKSREWYRANPNINRERRELMKWFLLKENEYLG 1006

Qy        584 --TELVKSLERN 593
                  | : |||::
Db       1007 QRMEKMDSLEKS 1018


RESULT 8
US-08-719-822B-2
; Sequence 2, Application US/08719822B
; Patent No. 5874527
;  GENERAL INFORMATION:
;    APPLICANT:  Barnwell, John
;    TITLE OF INVENTION:  Plasmodium vivax Blood Stage Antigens
;    NUMBER OF SEQUENCES:  4
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Darby and Darby
;      STREET:  805 Third Ave.
;      CITY:  New York
;      STATE:  New York
;      COUNTRY:  USA
;      ZIP:  10022-7513
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Floppy disk
;      COMPUTER:  IBM PC compatible
;      OPERATING SYSTEM:  PC-DOS/MS-DOS
;      SOFTWARE:  PatentIn Release #1.0, Version #1.25
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/719,822B
;      FILING DATE:  09/30/96
;      CLASSIFICATION:  435
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Gogoris, Adda
;      REGISTRATION NUMBER:  29,714
;      REFERENCE/DOCKET NUMBER:  5986/17686US2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (212)527-7700
;      TELEFAX:  (212)753-6237
;      TELEX:  236687
;  INFORMATION FOR SEQ ID NO:  2:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  1018 amino acids
;      TYPE:  amino acid
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
;    HYPOTHETICAL:  YES
;    ANTI-SENSE:  NO
;    FRAGMENT TYPE:  C-terminal
;    ORIGINAL SOURCE:
;      ORGANISM:  Plasmodium vivax
;    IMMEDIATE SOURCE:
;      CLONE:  PvMB3.3.1
US-08-719-822B-2

  Query Match             4.9%;  Score 197.5;  DB 2;  Length 1018;
  Best Local Similarity   23.2%;  Pred. No. 3.1e-07;
  Matches  128;  Conservative   86;  Mismatches  205;  Indels  133;  Gaps   27;

Qy        109 LKERSTSEPAVPPEEAEP-----HTEPEEQVPVEAEPQNIED--------EAKEQIQSLL 155
              ||:    |    | :  |         | ::|| |    |  ||          || |: : :|
Db        533 LKDPDAGEAVTVPSKEAPVQVPVAVGPAQEVPTEELMQLQEDDFELEGTAEAPEEGELVL 592

Qy        156 HEMVHAEHVEGEDLQQEDGP-TGEPQQED--DEFLMATDVDDRFETLEPEVSHEETEHSY 212
                  |||     |: |  ||| : :  :| | ||  || ||  ||  : || | :
Db        593 ---------EGEGEPTEEEPREGEPTEGEVPEEELEATPEDD-FELEEP--TGEEVEETV 640

Qy        213 HVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENE 272
                |||       :::||: :| |                  : ::|  :| | : | | :|
Db        641 EGEETAE---GEEVEEVPAEVE---------------EVEEVPAEVEEVEEVPEEVEEV 681
```

```
Qy        273 GIEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPK 332
Db        682 PAEVEEVEEVEPEE---------VEEVP----EEVEEVPEEVEEVPEEVEEVEEVEEVEVP 728

Qy        333 LLNKFDKTIKAELDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKR 392
Db        729 AVVEVEVPAVVEEEVPEEVEEEEEEEEPVEEEDVLQLVIPSEEDIQLDKPK-KDELGSGI 787

Qy        393 RS------NEVLRGAIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRL 446
Db        788 LSIIDMHYQDVPKEFMEEEEETAVYPLKPEDF----AKEDSQSTEWLTFIQGLEGDWERL 843

Qy        447 VQLFPNDTSL----------KNDLGVGYLLIGDNDNAKKVYEEVLSVTPND--GFAKVHY 494
Db        844 ------EVSLNKARERWMEQRNKEWAGWLRLIENKWSE--YSQI-STKGKDPAGLRKREW 894

Qy        495 G-------FILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGNKEAYKWY 547
Db        895 SDEKWKKWFKAEVKSQIDS---HLKKWMND----THSNLFKILVKD-MSQFENKKTKEWL 946

Qy        548 -------ELGHKRGHF-------------ASVWQRSLYNVNGLKAQ--PWWTPKETGY-- 583
Db        947 MNHWKKNERGYGSESFEVMTTSKLLNVAKSREWYRANPNINRERRELMKWFLLKENEYLG 1006

Qy        584 --TELVKSLERN 593
Db       1007 QRMEKMDSLEKS 1018
```

RESULT 9
US-09-092-458-2
; Sequence 2, Application US/09092458
; Patent No. 6231861
; GENERAL INFORMATION:
; APPLICANT: Barnwell, John
; TITLE OF INVENTION: Plasmodium vivax Blood Stage Antigens,
; TITLE OF INVENTION: Monoclonal Antibodies, and Diagnostic Assays
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Darby and Darby
; STREET: 805 Third Ave.
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10022-7513
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/092,458
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/719,821
; FILING DATE: 09/30/96
; ATTORNEY/AGENT INFORMATION:
; NAME: Gogoris, Adda
; REGISTRATION NUMBER: 29,714
; REFERENCE/DOCKET NUMBER: 5986/17686US3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212)527-7700
; TELEFAX: (212)753-6237
; TELEX: 236687
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1018 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ANTI-SENSE: NO
; FRAGMENT TYPE: C-terminal
; ORIGINAL SOURCE:
; ORGANISM: Plasmodium vivax
; IMMEDIATE SOURCE:
; CLONE: PvMB3.3.1
US-09-092-458-2

Query Match 4.9%; Score 197.5; DB 3; Length 1018;
Best Local Similarity 23.2%; Pred. No. 3.1e-07;
Matches 128; Conservative 86; Mismatches 205; Indels 133; Gaps 27;

```
Qy        109 LKERSTSEPAVPPEEAEP-----HTEPEEQVPVEAEPQNIED--------EAKEQIQSLL 155
Db        533 LKDPDAGEAVTVPSKEAPVQVPVAVGPAQEVPTEELMQLQEDDFELEGTAEAPEEGELVL 592

Qy        156 HEMVHAEHVEGEDLQQEDGP-TGEPQQED--DEFLMATDVDDRFETLEPEVSHEETEHSY 212
Db        593 ---------EGEGEPTEEEPREGEPTEGEVPEEELEATPEDD-FELEEP--TGEEVEETV 640

Qy        213 HVEETVSQDCNQDMEEMMSEQENPDSSEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENE 272
Db        641 EGEETAE---GEEVEEVPAEVE----------------EVEEVPAEVEEVEEVPEEVEEV 681

Qy        273 GIEITEVTAPPEDNPVEDSQVIVEEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPK 332
Db        682 PAEVEEVEEVEPEE---------VEEVP----EEVEEVPEEVEEVPEEVEEVEEVEEVEVP 728

Qy        333 LLNKFDKTIKAELDAAEKLRKRGKIEEAVNAFKELVRKYPQSPRARYGKAQCEDDLAEKR 392
Db        729 AVVEVEVPAVVEEEVPEEVEEEEEEEEPVEEEDVLQLVIPSEEDIQLDKPK-KDELGSGI 787

Qy        393 RS------NEVLRGAIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTLQRL 446
Db        788 LSIIDMHYQDVPKEFMEEEEETAVYPLKPEDF----AKEDSQSTEWLTFIQGLEGDWERL 843

Qy        447 VQLFPNDTSL----------KNDLGVGYLLIGDNDNAKKVYEEVLSVTPND--GFAKVHY 494
Db        844 ------EVSLNKARERWMEQRNKEWAGWLRLIENKWSE--YSQI-STKGKDPAGLRKREW 894

Qy        495 G-------FILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGNKEAYKWY 547
Db        895 SDEKWKKWFKAEVKSQIDS---HLKKWMND----THSNLFKILVKD-MSQFENKKTKEWL 946

Qy        548 -------ELGHKRGHF-------------ASVWQRSLYNVNGLKAQ--PWWTPKETGY-- 583
Db        947 MNHWKKNERGYGSESFEVMTTSKLLNVAKSREWYRANPNINRERRELMKWFLLKENEYLG 1006

Qy        584 --TELVKSLERN 593
Db       1007 QRMEKMDSLEKS 1018
```

RESULT 10
US-08-728-323A-2
; Sequence 2, Application US/08728323A
; Patent No. 5948676
; GENERAL INFORMATION:
; APPLICANT: Chang, Yuan
; APPLICANT: Bohenzky, Roy A.
; APPLICANT: Russo, James J.
; APPLICANT: Edelman, Isidore S.
; APPLICANT: Moore, Patrick S.
; TITLE OF INVENTION: Immediate Early Protein From Kaposi's
; TITLE OF INVENTION: Sarcoma-Associated Herpesvirus, DNA
; TITLE OF INVENTION: Encoding Same And Uses Thereof
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Cooper & Dunham LLP
; STREET: 1185 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.

```
;       ZIP:  10036
;     COMPUTER READABLE FORM:
;       MEDIUM TYPE:  Floppy disk
;       COMPUTER:  IBM PC compatible
;       OPERATING SYSTEM:  PC-DOS/MS-DOS
;       SOFTWARE:  PatentIn Release #1.0, Version #1.30
;     CURRENT APPLICATION DATA:
;       APPLICATION NUMBER:  US/08/728,323A
;       FILING DATE:
;       CLASSIFICATION:  435
;     ATTORNEY/AGENT INFORMATION:
;       NAME:  White, John P.
;       REGISTRATION NUMBER:  28,678
;       REFERENCE/DOCKET NUMBER:  0575/52268/JPW/MSC/SKS
;     TELECOMMUNICATION INFORMATION:
;       TELEPHONE:  212-278-0400
;       TELEFAX:  212-391-0525
;   INFORMATION FOR SEQ ID NO:  2:
;     SEQUENCE CHARACTERISTICS:
;      LENGTH: 1162 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
US-08-728-323A-2

  Query Match             4.7%;  Score 190.5;  DB 2;  Length 1162;
  Best Local Similarity   22.9%;  Pred. No. 1.4e-06;
  Matches   70;  Conservative   68;  Mismatches  123;  Indels   45;  Gaps   11;

Qy        122 EEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQ 181
              :| |   | |:|   | : :  :|| ::|          |  : :: ||:|    | ||
Db        698 DEQEQQDEQEQQDEQEQQDEQQQDEQQQQ----------DEQQQQDEQQQQD----EQQQ 743

Qy        182 EDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEM-----MSEQENP 236
              :|::       |  |  | :   || |     :|   :|  |::||        |||
Db        744 QDEQ-----QQQDEQEQQEEQEQQEEQEQELEEQEQELEDQEQELEEQEQELEEQEQELE 798

Qy        237 DSSEPVVEDER-LHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIV 295
              :  : : | |: |     ::  |  | :   : || :  |: |     ::  ||: :
Db        799 EQEQELEEQEQELEEQEQELEEQEQELEEQEQELEEQEQELEEQEVEEQEQEVEEQEQEQ 858

Qy        296 EEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRG 355
              ||  :  ||||     |  ::  : ::  :|:::: : | : ::  : ||:  |: :::
Db        859 EEQELEEVEEQ-----EQEQEEQEQELEEVEEQEEQELEEVEEQEEQELEEVEE-QEQQ 912

Qy        356 KIEEAVNAFKELVRKYPQ----SPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASL 411
              ::||     :: | :  |     |  :| :  ||::       ||    : |::::||
Db        913 ELEEVEEQEQQGVEQQEQETVEEPIILHGSSS-EDEM-------EVDYPVVSTHEQIASS 964

Qy        412 PDVPAD 417
              |  | |
Db        965 P--PGD 968


RESULT 11
US-09-298-568-2
; Sequence 2, Application US/09298568
; Patent No. 6322792
; GENERAL INFORMATION:
;  APPLICANT: Kieff, Elliott D.
;  APPLICANT: Ballestas, Mary E.
;  APPLICANT: Kaye, Kenneth M.
;  TITLE OF INVENTION: RHADINO VIRUS LANA ACTS IN TRANS ON A UNIT OF RHADINO
;  TITLE OF INVENTION: VIRUS DNA TO MEDIATE EFFICIENT EPISOME PERSISTENCE
;  FILE REFERENCE: 16412-10001R
;  CURRENT APPLICATION NUMBER: US/09/298,568
;  CURRENT FILING DATE: 1999-04-21
;  EARLIER APPLICATION NUMBER: US 60/109,422
;  EARLIER FILING DATE: 1998-11-19
;  NUMBER OF SEQ ID NOS: 3
;  SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-298-568-2

  Query Match             4.7%;  Score 190.5;  DB 4;  Length 1162;
  Best Local Similarity   22.9%;  Pred. No. 1.4e-06;
  Matches   70;  Conservative   68;  Mismatches  123;  Indels   45;  Gaps   11;

Qy        122 EEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQ 181
              :| |   | |:|   | : :  :|| ::|          |  : :: ||:|    | ||
Db        698 DEQEQQDEQEQQDEQEQQDEQQQDEQQQQ----------DEQQQQDEQQQQD----EQQQ 743

Qy        182 EDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEM-----MSEQENP 236
              :|::       |  |  | :   || |     :|   :|  |::||        |||
Db        744 QDEQ-----QQQDEQEQQEEQEQQEEQEQELEEQEQELEDQEQELEEQEQELEEQEQELE 798

Qy        237 DSSEPVVEDER-LHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIV 295
              :  : : | |: |     ::  |  | :   : || :  |: |     ::  ||: :
Db        799 EQEQELEEQEQELEEQEQELEEQEQELEEQEQELEEQEQELEEQEVEEQEQEVEEQEQEQ 858

Qy        296 EEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRG 355
              ||  :  ||||     |  ::  : ::  :|:::: : | : ::  : ||:  |: :::
Db        859 EEQELEEVEEQ-----EQEQEEQEQELEEVEEQEEQELEEVEEQEEQELEEVEE-QEQQ 912


RESULT 12
US-09-410-399-2
; Sequence 2, Application US/09410399
; Patent No. 6482587
; GENERAL INFORMATION:
;  APPLICANT:  Robertson, Erle S.
;  APPLICANT:  Cotter, Murray A.
;  TITLE OF INVENTION:  Methods to Inhibit or Enhance the Binding of Viral DNA
;  TITLE OF INVENTION:  to Genomic Host DNA
;  FILE REFERENCE:  UM-03778
;  CURRENT APPLICATION NUMBER: US/09/410,399
;  CURRENT FILING DATE:  1999-10-01
;  NUMBER OF SEQ ID NOS:  6
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 1162
;   TYPE: PRT
;   ORGANISM: Kaposi's sarcoma-associated herpesvirus
US-09-410-399-2

  Query Match             4.7%;  Score 190.5;  DB 4;  Length 1162;
  Best Local Similarity   22.9%;  Pred. No. 1.4e-06;
  Matches   70;  Conservative   68;  Mismatches  123;  Indels   45;  Gaps   11;

Qy        122 EEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQQ 181
              :| |   | |:|   | : :  :|| ::|          |  : :: ||:|    | ||
Db        698 DEQEQQDEQEQQDEQEQQDEQQQDEQQQQ----------DEQQQQDEQQQQD----EQQQ 743

Qy        182 EDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEM-----MSEQENP 236
              :|::       |  |  | :   || |     :|   :|  |::||        |||
Db        744 QDEQ-----QQQDEQEQQEEQEQQEEQEQELEEQEQELEDQEQELEEQEQELEEQEQELE 798

Qy        237 DSSEPVVEDER-LHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIV 295
              :  : : | |: |     ::  |  | :   : || :  |: |     ::  ||: :
Db        799 EQEQELEEQEQELEEQEQELEEQEQELEEQEQELEEQEQELEEQEVEEQEQEVEEQEQEQ 858

Qy        296 EEVSIFPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRG 355
              ||  :  ||||     |  ::  : ::  :|:::: : | : ::  : ||:  |: :::
Db        859 EEQELEEVEEQ-----EQEQEEQEQELEEVEEQEEQELEEVEE-QEQQ 912
```

```
Qy        356 KIEEAVNAFKELVRKYPQ----SPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASL 411
Db        913 ELEEVEEQEQQGVEQQEQETVEEPIILHGSSS-EDEM-------EVDYPVVSTHEQIASS 964

Qy        412 PDVPAD 417
Db        965 P--PGD 968


RESULT 13
US-09-328-352-8015
; Sequence 8015, Application US/09328352
; Patent No. 6562958
; GENERAL INFORMATION:
;  APPLICANT: Gary L. Breton et al.
;  TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ACINETOBACTER
;  TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS
;  FILE REFERENCE: GTC99-03PA
;  CURRENT APPLICATION NUMBER: US/09/328,352
;  CURRENT FILING DATE: 1999-06-04
;  NUMBER OF SEQ ID NOS: 8252
; SEQ ID NO 8015
;   LENGTH: 312
;   TYPE: PRT
;   ORGANISM: Acinetobacter baumannii
US-09-328-352-8015

  Query Match             4.7%;  Score 189.5;  DB 4;  Length 312;
  Best Local Similarity   27.8%;  Pred. No. 2.5e-07;
  Matches   52;  Conservative   34;  Mismatches   88;  Indels   13;  Gaps    4;

Qy        564 LYNVVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDEGLAVMD----KAKGLFLPEDENL 619
Db         53 MYIFSKVPNQPY---IDTQHFKDLKVLDENWEMIRDEAKALYDQGGIKASSTYNDLGFNS 109

Qy        620 REKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGCRRGQIKYSIMHPGTHVWPH 679
Db        110 FFKTGWKRFYLKWYESSHPSAAELCPKTTALLKTLPTIKAA-----MFTELAPDSRLVRH 164

Qy        680 TGPTNCRLRMHLGLVIPKEG-CKIRCANETRTWEEGKVLIFDDSFEHEVWQDASSFRLIF 738
Db        165 RDPYAGSLRYHLGLITPNDDRCFIDVDGERYSWRDGQSVVFDETYIHYAENKTDQNRIIF 224

Qy        739 IVDVWHP 745
Db        225 FADVERP 231


RESULT 14
US-08-961-083-218
; Sequence 218, Application US/08961083
; Patent No. 6159469
;  GENERAL INFORMATION:
;    APPLICANT:  Choi et. al.
;    TITLE OF INVENTION:  Streptococcus pneumoniae Antigens and Vaccines
;    NUMBER OF SEQUENCES:  452
;    CORRESPONDENCE ADDRESS:
;      ADDRESSEE:  Human Genome Sciences, Inc.
;      STREET:  9410 Key West Avenue
;      CITY:  Rockville
;      STATE:  Maryland
;      COUNTRY:  USA
;      ZIP:  20850
;    COMPUTER READABLE FORM:
;      MEDIUM TYPE:  Diskette, 3.50 inch, 1.4Mb storage
;      COMPUTER:  HP Vectra 486/33
;      OPERATING SYSTEM:  MSDOS version 6.2
;      SOFTWARE:  ASCII Text
;    CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/08/961,083
;      FILING DATE:
;      CLASSIFICATION:  435
;    PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;    ATTORNEY/AGENT INFORMATION:
;      NAME:  Brookes, A. Anders
;      REGISTRATION NUMBER:  36,373
;      REFERENCE/DOCKET NUMBER:  PB340P2
;    TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  (301) 309-8504
;      TELEFAX:  (301) 309-8512
;  INFORMATION FOR SEQ ID NO:  218:
;    SEQUENCE CHARACTERISTICS:
;      LENGTH:  565 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;    MOLECULE TYPE:  protein
US-08-961-083-218

  Query Match             4.4%;  Score 177.5;  DB 3;  Length 565;
  Best Local Similarity   22.5%;  Pred. No. 5.5e-06;
  Matches  114;  Conservative   71;  Mismatches  199;  Indels  123;  Gaps   22;

Qy        111 ERSTSEPAVPPEE--AEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGED 168
Db         62 KEDSAEPA-PVEEVGGEVESKPEEKVAVKPESQPSDKPAEESKVEQAGEPV----APRED 116

Qy        169 LQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQ-DME 227
Db        117 EKAPVEPEKQPEAPEEE--KAVEETPKQEESTPDTKAEETVEP--KEETVNQSIEQPKVE 172

Qy        228 EMMSEQENPDSSEPVVEDE----RLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPP 283
Db        173 TPAVEKQTEPTEEPKVEQAGEPVAPREDEQAPTAPVEPEKQPEVPEEEKAV---EETPKP 229

Qy        284 ED--------NPVEDSQV---IVEEVSIFPVE------------------------EQQ 307
Db        230 EDKIKGIGTKEPVDKSELNNQIDKASSVSPTDYSTASYNALGPVLETAKGVYASEPVKQP 289

Qy        308 EVPPETNR-----------KTD------DPEQKAKVKKKKPKLLNKFDKTIKAELDAAEK 350
Db        290 EVNSETNKLKTAIDALNVDKTELNNTIADAKTKVKEHYSDRSWQNLQTEVTKAEKVAANT 349

Qy        351 LRKRGKIEEAVNAF-------------------------KELVRKYPQSPRARYGKAQ 383
Db        350 DAKQSEVNEAVEKLTATIEKLVELSEKPILTLTSTDKKILEREAVAKYTLE---NQNKTK 406

Qy        384 CEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTL 443
Db        407 IKSITAELKKGEEVINTVVLTDDKVTT-ETISAAFKNLEYYKEYTLSTTMIYDRGNGEET 465

Qy        444 QRL----VQLFPNDTSLKN----DLGVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYG 495
Db        466 ETLENQNIQLDLKKVELKNIKRTDL-IKY------ENGKETNESLITTIPDD---KSNYY 515

Qy        496 FILKAQNKIAE--SIPYLKEGIESGDP 520
Db        516 LKITSNNQKTTLLAVKNIEETTVNGTP 542


RESULT 15
US-09-536-784-218
; Sequence 218, Application US/09536784
; Patent No. 6573082
;  GENERAL INFORMATION:
;       APPLICANT: Choi et. al.
;       TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
;       NUMBER OF SEQUENCES: 452
;       CORRESPONDENCE ADDRESS:
;            ADDRESSEE: Human Genome Sciences, Inc.
;            STREET: 9410 Key West Avenue
;            CITY: Rockville
;            STATE: Maryland
;            COUNTRY: USA
```

```
;            ZIP: 20850
;        COMPUTER READABLE FORM:
;            MEDIUM TYPE: Diskette, 3.50 inch, 1.4Mb storage
;            COMPUTER: HP Vectra 486/33
;            OPERATING SYSTEM: MSDOS version 6.2
;            SOFTWARE: ASCII Text
;        CURRENT APPLICATION DATA:
;            APPLICATION NUMBER: US/09/536,784
;            FILING DATE: 30-Oct-1997
;            CLASSIFICATION: <Unknown>
;        PRIOR APPLICATION DATA:
;            APPLICATION NUMBER: 08/961,083
;            FILING DATE: OCT-30-1997
;        ATTORNEY/AGENT INFORMATION:
;            NAME: Michelle S. Marks
;            REGISTRATION NUMBER: 41,971
;            REFERENCE/DOCKET NUMBER: PB340P3
;        TELECOMMUNICATION INFORMATION:
;            TELEPHONE: (301) 309-8504
;            TELEFAX: (301) 309-8512
;   INFORMATION FOR SEQ ID NO: 218:
;        SEQUENCE CHARACTERISTICS:
;            LENGTH: 565 amino acids
;            TYPE: amino acid
;            STRANDEDNESS: single
;            TOPOLOGY: linear
;        MOLECULE TYPE: protein
;        SEQUENCE DESCRIPTION: SEQ ID NO: 218:
US-09-536-784-218

  Query Match             4.4%;  Score 177.5;  DB 4;  Length 565;
  Best Local Similarity   22.5%;  Pred. No. 5.5e-06;
  Matches  114;  Conservative   71;  Mismatches  199;  Indels  123;  Gaps   22;

Qy        111 ERSTSEPAVPPEE--AEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGED 168
              :  ::||| | ||   |  ::|||:| |: | |  :  |:|      | |       ||
Db         62 KEDSAEPA-PVEEVGGEVESKPEEKVAVKPESQPSDKPAEESKVEQAGEPV----APRED 116

Qy        169 LQQEDGPTGEPQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQ-DME 227
               :    |  :|:  ::|   | :   : |   |:   |||     ||||:|   |  :|
Db        117 EKAPVEPEKQPEAPEEE--KAVEETPKQEESTPDTKAEETVEP--KEETVNQSIEQPKVE 172

Qy        228 EMMSEQENPDSSEPVVEDE----RLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPP 283
                 |::   : || ||          |    | |  |:    | | : :   | |  |
Db        173 TPAVEKQTEPTEEPKVEQAGEPVAPREDEQAPTAPVEPEKQPEVPEEEKAV---EETPKP 229

Qy        284 ED--------NPVEDSQV---IVEEVSIFPVE-------------------------EQQ 307
              ||          ||: |::   | :  |: | :                        :|
Db        230 EDKIKGIGTKEPVDKSELNNQIDKASSVSPTDYSTASYNALGPVLETAKGVYASEPVKQP 289

Qy        308 EVPPETNR-----------KTD------DPEQKAKVKKKKPKLLNKFDKTIKAELDAAEK 350
              ||  |||:           ||:       | : | |          |   :  |||  ||
Db        290 EVNSETNKLKTAIDALNVDKTELNNTIADAKTKVKEHYSDRSWQNLQTEVTKAEKVAANT 349

Qy        351 LRKRGKIEEAVNAF--------------------------KELVRKYPQSPRARYGKAQ 383
                |: :: |||                          :| | ||          | :
Db        350 DAKQSEVNEAVEKLTATIEKLVELSEKPILTLTSTDKKILEREAVAKYTLE---NQNKTK 406

Qy        384 CEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLKLSLKRRSDRQQFLGHMRGSLLTL 443
               :   || ::  ||:   : |  :| :   : |    |   :       : : ||:
Db        407 IKSITAELKKGEEVINTVVLTDDKVTT-ETISAAFKNLEYYKEYTLSTTMIYDRGNGEET 465

Qy        444 QRL----VQLFPNDTSLKN----DLGVGYLLIGDNDNAKKVYEEVLSVTPNDGFAKVHYG 495
               : |    :||      |||    || : |       :| |:  | :::  |:|   | :|
Db        466 ETLENQNIQLDLKKVELKNIKRTDL-IKY------ENGKETNESLITTIPDD---KSNYY 515

Qy        496 FILKAQNKIAE--SIPYLKEGIESGDP 520
               : : |:     ::  ::|   :| |
Db        516 LKITSNNQKTTLLAVKNIEETTVNGTP 542


Search completed: May  5, 2004, 11:05:05
Job time : 25 secs
```

GenCore version 5.1.6

OM protein - protein search, using sw model

Run on: May 5, 2004, 11:03:59 ; Search time 48 Seconds
(without alignments)
4377.285 Million cell updates/sec

Title: US-09-903-216-2
Perfect score: 4022
Sequence: 1 MAQRKNAKSSGNSSSSGSGS.....IVDVWHPELTPQQRRSLPAI 758

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1138120 seqs, 277189581 residues

Total number of hits satisfying chosen parameters: 1138120

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB	ID	Description
1	4022	100.0	758	9	US-09-903-248-2	Sequence 2, Appli
2	4022	100.0	758	9	US-09-859-604-2	Sequence 2, Appli
3	4022	100.0	758	9	US-09-903-063-2	Sequence 2, Appli
4	4022	100.0	758	9	US-09-903-216-2	Sequence 2, Appli
5	4022	100.0	758	9	US-09-903-199-2	Sequence 2, Appli
6	4022	100.0	758	9	US-09-903-023-2	Sequence 2, Appli
7	4022	100.0	758	10	US-09-436-184-2	Sequence 2, Appli
8	1320	32.8	255	9	US-09-736-457-1806	Sequence 1806, Ap
9	1320	32.8	255	9	US-09-902-941-1806	Sequence 1806, Ap
10	1320	32.8	255	9	US-09-849-626-1806	Sequence 1806, Ap
11	1320	32.8	255	12	US-10-283-017-1806	Sequence 1806, Ap
12	1320	32.8	255	14	US-10-017-754-1806	Sequence 1806, Ap
13	1320	32.8	255	14	US-10-113-872-1806	Sequence 1806, Ap
14	275	6.8	250	14	US-10-156-761-8153	Sequence 8153, Ap
15	274.5	6.8	369	14	US-10-168-274-15	Sequence 15, Appl
16	218.5	5.4	299	12	US-10-282-122A-49023	Sequence 49023, A
17	217.5	5.4	324	12	US-10-221-278-250	Sequence 250, App
18	217.5	5.4	324	15	US-10-291-172-250	Sequence 250, App
19	199	4.9	109	14	US-10-168-274-5	Sequence 5, Appli
20	196.5	4.9	355	12	US-10-282-122A-51120	Sequence 51120, A
21	190.5	4.7	1162	12	US-09-894-273-2	Sequence 2, Appli
22	190.5	4.7	1162	14	US-10-294-804-2	Sequence 2, Appli
23	189.5	4.7	304	12	US-10-282-122A-45188	Sequence 45188, A
24	189.5	4.7	1616	10	US-09-820-843A-16	Sequence 16, Appl
25	189.5	4.7	1616	12	US-10-282-122A-63593	Sequence 63593, A
26	186	4.6	50	16	US-10-443-622-48	Sequence 48, Appl
27	182	4.5	350	13	US-10-101-487-58	Sequence 58, Appl
28	177.5	4.4	565	9	US-09-765-272-218	Sequence 218, App
29	177	4.4	240	13	US-10-101-487-75	Sequence 75, Appl
30	177	4.4	845	12	US-10-205-331-61	Sequence 61, Appl
31	175	4.4	1000	14	US-10-128-714-3305	Sequence 3305, Ap
32	174	4.3	566	12	US-10-424-599-172054	Sequence 172054,
33	172.5	4.3	1302	12	US-10-282-122A-46112	Sequence 46112, A
34	170	4.2	299	12	US-10-282-122A-50678	Sequence 50678, A
35	168.5	4.2	200	13	US-10-101-487-53	Sequence 53, Appl
36	168.5	4.2	300	12	US-10-282-122A-60256	Sequence 60256, A
37	168	4.2	302	12	US-10-282-122A-75844	Sequence 75844, A
38	167.5	4.2	1876	12	US-10-282-122A-73861	Sequence 73861, A
39	167	4.2	187	13	US-10-101-487-50	Sequence 50, Appl
40	167	4.2	302	9	US-09-815-242-13815	Sequence 13815, A
41	165.5	4.1	1095	14	US-10-128-714-8305	Sequence 8305, Ap
42	165	4.1	518	12	US-10-425-114-71472	Sequence 71472, A
43	164	4.1	1957	15	US-10-369-493-2070	Sequence 2070, Ap
44	163.5	4.1	1297	12	US-10-282-122A-71692	Sequence 71692, A
45	163.5	4.1	1501	9	US-09-924-154-17	Sequence 17, Appl

ALIGNMENTS

RESULT 1
US-09-903-248-2
; Sequence 2, Application US/09903248
; Patent No. US20020102263A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV5
; CURRENT APPLICATION NUMBER: US/09/903,248
; CURRENT FILING DATE: 2001-07-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-248-2

Query Match 100.0%; Score 4022; DB 9; Length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
```

```
Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||
Db        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
```

RESULT 2
US-09-859-604-2
; Sequence 2, Application US/09859604
; Patent No. US20020110559A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M
; APPLICANT: Deutch, Alan H
; APPLICANT: Ghanbari, Hossein A
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 CIP
; CURRENT APPLICATION NUMBER: US/09/859,604
; CURRENT FILING DATE: 2001-05-17
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-859-604-2

Query Match 100.0%; Score 4022; DB 9; Length 758;
Best Local Similarity 100.0%; Pred. No. 1.5e-282;
Matches 758; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
```

```
Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||
Db        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
```

RESULT 3
US-09-903-063-2
; Sequence 2, Application US/09903063
; Patent No. US20020114810A1
; GENERAL INFORMATION:
; APPLICANT: Wands, Jack R.
; APPLICANT: de la Monte, Suzanne M.
; APPLICANT: Ince, Nedim
; APPLICANT: Carlson, Rolf I.
; TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
; FILE REFERENCE: 21486-032 DIV3
; CURRENT APPLICATION NUMBER: US/09/903,063
; CURRENT FILING DATE: 2001-10-11
; PRIOR APPLICATION NUMBER: 09/436,184
; PRIOR FILING DATE: 1999-11-08
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 758
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-903-063-2

Query Match 100.0%; Score 4022; DB 9; Length 758;

```
  Best Local Similarity   100.0%;  Pred. No. 1.5e-282;
  Matches  758;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||
Db        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758


RESULT 4
US-09-903-216-2
; Sequence 2, Application US/09903216
; Patent No. US20020114811A1
; GENERAL INFORMATION:
;  APPLICANT: Wands, Jack R.
;  APPLICANT:  de la Monte, Suzanne M.
;  APPLICANT:  Ince, Nedim
;  APPLICANT:  Carlson, Rolf I.
;  TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
;  FILE REFERENCE: 21486-032 DIV2
;  CURRENT APPLICATION NUMBER: US/09/903,216
;  CURRENT FILING DATE:  2001-07-11
;  PRIOR APPLICATION NUMBER: 09/436,184
;  PRIOR FILING DATE: 1999-11-08
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1

; SEQ ID NO 2
;   LENGTH: 758
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-903-216-2

  Query Match           100.0%;  Score 4022;  DB 9;  Length 758;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-282;
  Matches  758;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||
Db        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758


RESULT 5
US-09-903-199-2
; Sequence 2, Application US/09903199
; Patent No. US20020122802A1
; GENERAL INFORMATION:
;  APPLICANT: Wands, Jack R.
;  APPLICANT:  de la Monte, Suzanne M.
;  APPLICANT:  Ince, Nedim
;  APPLICANT:  Carlson, Rolf I.
;  TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
```

```
;  FILE REFERENCE: 21486-032 DIV4
;  CURRENT APPLICATION NUMBER: US/09/903,199
;  CURRENT FILING DATE:  2001-07-11
;  PRIOR APPLICATION NUMBER: 09/436,184
;  PRIOR FILING DATE: 1999-11-08
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 758
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-903-199-2

  Query Match             100.0%;  Score 4022;  DB 9;  Length 758;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-282;
  Matches  758;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||||
Db        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758


RESULT 6
US-09-903-023-2
; Sequence 2, Application US/09903023
; Patent No. US20020146421A1
; GENERAL INFORMATION:
;  APPLICANT: Wands, Jack R.
;  APPLICANT:  de la Monte, Suzanne M.
;  APPLICANT:  Ince, Nedim
;  APPLICANT:  Carlson, Rolf I.
;  TITLE OF INVENTION: DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
;  FILE REFERENCE: 21486-032 DIV1
;  CURRENT APPLICATION NUMBER: US/09/903,023
;  CURRENT FILING DATE:  2001-10-11
;  PRIOR APPLICATION NUMBER: 09/436,184
;  PRIOR FILING DATE: 1999-11-08
;  NUMBER OF SEQ ID NOS: 9
;  SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
;   LENGTH: 758
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-903-023-2

  Query Match             100.0%;  Score 4022;  DB 9;  Length 758;
  Best Local Similarity   100.0%;  Pred. No. 1.5e-282;
  Matches  758;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
```

```
                                     ||||||||||||||||||||||||||||||||||||
Db        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758


RESULT 7
US-09-436-184-2
; Sequence 2, Application US/09436184
; Publication No. US20030031670A1
; GENERAL INFORMATION:
;  APPLICANT:  Wands, Jack R.
;  APPLICANT:  de la Monte, Suzanne M.
;  APPLICANT:  Ince, Nedim
;  APPLICANT:  Carlson, Rolf I.
;  TITLE OF INVENTION:  DIAGNOSIS AND TREATMENT OF MALIGNANT NEOPLASMS
;  FILE REFERENCE:  R.I. Hosp. - Malignant Neoplasms
;  CURRENT APPLICATION NUMBER: US/09/436,184
;  CURRENT FILING DATE:  1999-11-08
;  NUMBER OF SEQ ID NOS:  7
;  SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
;   LENGTH: 758
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-436-184-2

  Query Match           100.0%;  Score 4022;  DB 10;  Length 758;
  Best Local Similarity  100.0%;  Pred. No. 1.5e-282;
  Matches  758;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MAQRKNAKSSGNSSSSGSGSGSTSAGSSSPGARRETKHGGHKNGRKGGLSGTSFFTWFMV 60

Qy         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 IALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPAVP 120

Qy        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 PEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGEPQ 180

Qy        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 QEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDSSE 240

Qy        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 PVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEVSI 300

Qy        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 FPVEEQQEVPPETNRKTDDPEQKAKVKKKKPKLLNKFDKTIKAELDAAEKLRKRGKIEEA 360

Qy        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 VNAFKELVRKYPQSPRARYGKAQCEDDLAEKRRSNEVLRGAIETYQEVASLPDVPADLLK 420

Qy        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 LSLKRRSDRQQFLGHMRGSLLTLQRLVQLFPNDTSLKNDLGVGYLLIGDNDNAKKVYEEV 480

Qy        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 LSVTPNDGFAKVHYGFILKAQNKIAESIPYLKEGIESGDPGTDDGRFYFHLGDAMQRVGN 540

Qy        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 KEAYKWYELGHKRGHFASVWQRSLYNVNGLKAQPWWTPKETGYTELVKSLERNWKLIRDE 600

Qy        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 GLAVMDKAKGLFLPEDENLREKGDWSQFTLWQQGRRNENACKGAPKTCTLLEKFPETTGC 660

Qy        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 RRGQIKYSIMHPGTHVWPHTGPTNCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFD 720

Qy        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758
              |||||||||||||||||||||||||||||||||||||
Db        721 DSFEHEVWQDASSFRLIFIVDVWHPELTPQQRRSLPAI 758


RESULT 8
US-09-736-457-1806
; Sequence 1806, Application US/09736457
; Patent No. US20020168637A1
; GENERAL INFORMATION:
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  Lodes, Michael A.
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Vedvick, Tom
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Retter, Marc
;  APPLICANT:  Mannion, Jane
;  APPLICANT:  Fan, Liqun
;  APPLICANT:  Wang, Aijun
;  TITLE OF INVENTION:  COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE:  210121.478C15
;  CURRENT APPLICATION NUMBER: US/09/736,457
;  CURRENT FILING DATE:  2000-12-13
;  NUMBER OF SEQ ID NOS:  1864
;  SOFTWARE:  FastSEQ for Windows Version 3.0
; SEQ ID NO 1806
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-736-457-1806

  Query Match           32.8%;  Score 1320;  DB 9;  Length 255;
  Best Local Similarity  98.8%;  Pred. No. 1.9e-87;
  Matches  252;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy        179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
              |||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||
Db        121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy        239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy        299 SIFPVEEQQEVPPET 313
              |||||||||||||:|
Db        241 SIFPVEEQQEVPPDT 255


RESULT 9
US-09-902-941-1806
; Sequence 1806, Application US/09902941
; Patent No. US20020172952A1
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Johnson, Jeffrey C.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Marnerakis, Margarita
```

```
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  McNabb, Andria
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C17
;  CURRENT APPLICATION NUMBER: US/09/902,941
;  CURRENT FILING DATE:  2001-07-10
;  NUMBER OF SEQ ID NOS: 2002
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1806
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-902-941-1806

  Query Match             32.8%;  Score 1320;  DB 9;  Length 255;
  Best Local Similarity   98.8%;  Pred. No. 1.9e-87;
  Matches  252;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy        179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy        239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy        299 SIFPVEEQQEVPPET 313
              |||||||||||||:|
Db        241 SIFPVEEQQEVPPDT 255


RESULT 10
US-09-849-626-1806
; Sequence 1806, Application US/09849626
; Publication No. US20020197669A1
; GENERAL INFORMATION:
;  APPLICANT: Bangur, Chaitanya
;  APPLICANT:  Fanger, Gary
;  APPLICANT:  Wang, Aijun
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Switzer, Anne
;  APPLICANT:  McNeill, Patricia
;  APPLICANT:  Clapper, Jonathan
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
;  TITLE OF INVENTION:  DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C16
;  CURRENT APPLICATION NUMBER: US/09/849,626
;  CURRENT FILING DATE:  2001-05-03
;  NUMBER OF SEQ ID NOS: 1926
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1806
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-09-849-626-1806

  Query Match             32.8%;  Score 1320;  DB 9;  Length 255;
  Best Local Similarity   98.8%;  Pred. No. 1.9e-87;
  Matches  252;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy        179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy        239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy        299 SIFPVEEQQEVPPET 313
              |||||||||||||:|
Db        241 SIFPVEEQQEVPPDT 255


RESULT 11
US-10-283-017-1806
; Sequence 1806, Application US/10283017
; Publication No. US20030211510A1
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Sleath, Paul R.
;  APPLICANT:  Johnson, Jeffrey C.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Durham, Margarita
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  McNabb, Andria
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C20
;  CURRENT APPLICATION NUMBER: US/10/283,017
;  CURRENT FILING DATE:  2002-10-28
;  NUMBER OF SEQ ID NOS: 2157
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1806
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-283-017-1806

  Query Match             32.8%;  Score 1320;  DB 12;  Length 255;
  Best Local Similarity   98.8%;  Pred. No. 1.9e-87;
  Matches  252;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy        179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy        239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy        299 SIFPVEEQQEVPPET 313
              |||||||||||||:|
```

```
Db        241 SIFPVEEQQEVPPDT 255


RESULT 12
US-10-017-754-1806
; Sequence 1806, Application US/10017754
; Publication No. US20030054363A1
; GENERAL INFORMATION:
;  APPLICANT: Henderson, Robert A.
;  APPLICANT:  Wang, Tongtong
;  APPLICANT:  Watanabe, Yoshihiro
;  APPLICANT:  Johnson, Jeffrey C.
;  APPLICANT:  Retter, Marc W.
;  APPLICANT:  Marnerakis, Margarita
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Bangur, Chaitanya S.
;  APPLICANT:  McNabb, Andria
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C18
;  CURRENT APPLICATION NUMBER: US/10/017,754
;  CURRENT FILING DATE:  2001-10-29
;  NUMBER OF SEQ ID NOS: 2004
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1806
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-017-754-1806

  Query Match           32.8%;  Score 1320;  DB 14;  Length 255;
  Best Local Similarity   98.8%;  Pred. No. 1.9e-87;
  Matches  252;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy        179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy        239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy        299 SIFPVEEQQEVPPET 313
              |||||||||||||:|
Db        241 SIFPVEEQQEVPPDT 255


RESULT 13
US-10-113-872-1806
; Sequence 1806, Application US/10113872
; Publication No. US20030170255A1
; GENERAL INFORMATION:
;  APPLICANT: Watanabe, Yoshihiro
;  APPLICANT:  Henderson, Robert A.
;  APPLICANT:  Kalos, Michael D.
;  APPLICANT:  Sleath, Paul R.
;  APPLICANT:  Vedvick, Thomas S.
;  APPLICANT:  Carter, Darrick
;  APPLICANT:  Fanger, Gary R.
;  TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
;  TITLE OF INVENTION:  AND DIAGNOSIS OF LUNG CANCER
;  FILE REFERENCE: 210121.478C19
;  CURRENT APPLICATION NUMBER: US/10/113,872
;  CURRENT FILING DATE:  2002-03-28
;  NUMBER OF SEQ ID NOS: 2011
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1806
;   LENGTH: 255
;   TYPE: PRT
;   ORGANISM: Homo sapiens
US-10-113-872-1806

  Query Match           32.8%;  Score 1320;  DB 14;  Length 255;
  Best Local Similarity   98.8%;  Pred. No. 1.9e-87;
  Matches  252;  Conservative    1;  Mismatches    2;  Indels    0;  Gaps    0;

Qy         59 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 118
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 MVIALLGVWTSVAVVWFDLVDYEEVLGKLGIYDADGDGDFDVDDAKVLLGLKERSTSEPA 60

Qy        119 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 178
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 VPPEEAEPHTEPEEQVPVEAEPQNIEDEAKEQIQSLLHEMVHAEHVEGEDLQQEDGPTGE 120

Qy        179 PQQEDDEFLMATDVDDRFETLEPEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 238
              |||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        121 PQQEDDEFLMATDVDDRFETLELEVSHEETEHSYHVEETVSQDCNQDMEEMMSEQENPDS 180

Qy        239 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTAPPEDNPVEDSQVIVEEV 298
              |||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||||
Db        181 SEPVVEDERLHHDTDDVTYQVYEEQAVYEPLENEGIEITEVTVPPEDNPVEDSQVIVEEV 240

Qy        299 SIFPVEEQQEVPPET 313
              |||||||||||||:|
Db        241 SIFPVEEQQEVPPDT 255


RESULT 14
US-10-156-761-8153
; Sequence 8153, Application US/10156761
; Publication No. US20030119018A1
; GENERAL INFORMATION:
;  APPLICANT: OMURA, SATOSHI
;  APPLICANT:  IKEDA, HARUO
;  APPLICANT:  ISHIKAWA, JUN
;  APPLICANT:  HORIKAWA, HIROSHI
;  APPLICANT:  SHIBA, TADAYOSHI
;  APPLICANT:  SAKAKI, YOSHIYUKI
;  APPLICANT:  HATTORI, MASAHIRA
;  TITLE OF INVENTION: NOVEL POLYNUCLEOTIDES
;  FILE REFERENCE: 249-262
;  CURRENT APPLICATION NUMBER: US/10/156,761
;  CURRENT FILING DATE:  2002-05-29
;  PRIOR APPLICATION NUMBER: JP 2001-204089
;  PRIOR FILING DATE: 2001-05-30
;  PRIOR APPLICATION NUMBER: JP 2001-272697
;  PRIOR FILING DATE: 2001-08-02
;  NUMBER OF SEQ ID NOS: 15109
; SEQ ID NO 8153
;   LENGTH: 250
;   TYPE: PRT
;   ORGANISM: Streptomyces avermitilis
US-10-156-761-8153

  Query Match           6.8%;  Score 275;  DB 14;  Length 250;
  Best Local Similarity   33.3%;  Pred. No. 9.2e-12;
  Matches   65;  Conservative   31;  Mismatches   91;  Indels    8;  Gaps    4;

Qy        569 GLKAQPWWTPKETGYTEL---VKSLERNWKLIRDEGLAVMDKAKGLFLPEDENLREKGDW 625
              ||  ||  |    | ||   |  || : : |::|       :  |   :  |  : :|
Db         48 GLSPTPWHDP--YAYAELAPVVHELEASHQAIKEELKVAWSARREAFSDYEHYLTRQDNW 105

Qy        626 SQFTLWQQGRRNENACKGAPKTCTLLEKFPETTG--CRRGQIKYSIMHPGTHVWPHTGPT 683
                |:::|   | :    |   :|:    ||  |   :  :| : ||  : ||
Db        106 QALYLFREGGITEESAATVPIAYQVLKDVAVDTGKICPLLECHFSTLLPGAVIEPHCDLW 165
```

```
Qy        684 NCRLRMHLGLVIPKEGCKIRCANETRTWEEGKVLIFDDSFEHEVWQDASSFRLIFIVDVW 743
              |  : :|| : || :|| |  | |||:||||| |:|| |||||     :  |   ::|:|
Db        166 NFSINLHLAVDIP-DGCSITVAGETRSWEEGKCLLFDYSFEHEARNTGTRPRTCLLIDLW 224

Qy        744 HPELTPQQRRSLPAI 758
              ||| |  :||:| |:
Db        225 HPETTVPERRALVAL 239


RESULT 15
US-10-168-274-15
; Sequence 15, Application US/10168274
; Publication No. US20030124106A1
; GENERAL INFORMATION:
;  APPLICANT: INCYTE GENOMICS, INC.
;  APPLICANT:  YUE, Henry
;  APPLICANT:  LAL, Preeti
;  APPLICANT:  TANG, Y. Tom
;  APPLICANT:  HILLMAN, Jennifer
;  APPLICANT:  BAUGHN, Mariah R.
;  APPLICANT:  AZIMZAI, Yalda
;  APPLICANT:  LU, Dyung Aina M.
;  TITLE OF INVENTION: HUMAN OXIDOREDUCTASE PROTEINS
;  FILE REFERENCE: PF-0754 PCT
;  CURRENT APPLICATION NUMBER: US/10/168,274
;  CURRENT FILING DATE:  2002-08-26
;  PRIOR APPLICATION NUMBER: 60/172,367
;  PRIOR FILING DATE: 1999-12-16
;  NUMBER OF SEQ ID NOS: 54
;  SOFTWARE: PERL Program
; SEQ ID NO 15
;   LENGTH: 369
;   TYPE: PRT
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: misc_feature
;   OTHER INFORMATION: Incyte ID No. US20030124106A1 2754425CD1
US-10-168-274-15

  Query Match             6.8%;  Score 274.5;  DB 14;  Length 369;
  Best Local Similarity   30.1%;  Pred. No. 1.7e-11;
  Matches   75;  Conservative   44;  Mismatches  105;  Indels   25;  Gaps    9;

Qy        526 RFYFHLGDAMQRVGNKEAYKWYELG--HK----RGHF----ASVWQRSLYNVNGLKAQPW 575
              : | :| :  :|      | |  :|  ||    :| :     |: :  :: :  |   |:
Db        125 KLYHNLQEYAKR------YSWSGMGRIHKGIREQGRYLNSRPSIQKPEVFFLPDLPTTPY 178

Qy        576 WTPKETGYTELVKSLERNWKLIRDEGLAVMDKAKGLFLPE--DENLREKGDWSQFTLWQQ 633
              ::     :   |: ||||:: |  |   :        ||:   |    |:|  | |  |
Db        179 FSRDAQKHD--VEVLERNFQTILCEFETLYKAFSNCSLPQGWKMNSTPSGEWFTFYLVNQ 236

Qy        634 GRRNENACKGAPKTCTLLEKFPETTGCR-RGQIKYSIMHPGTHVWPHTGPTNCRLRMHLG 692
              |       |:  |:|  ||        |     |    |:: ||| :  | |||| |:| |||
Db        237 GVCVPRNCRKCPRTYRLLGSLRTCIGNNVFGNACISVLSPGTVITEHYGPTNIRIRCHLG 296

Qy        693 LVIPKEGCKIRCANETRTWEEGKVLIFDDSFEHEVWQDASSF---RLIFIVDVWHPELTP 749
              |  |  ||::    | : | ||: |:|||||  |  : : |:    |::|:||:||| :
Db        297 LKTP-NGCELVVGGEPQCWAEGRCLLFDDSFLHAAFHEGSAEDGPRVVFMVDLWHPNVAA 355

Qy        750 QQRRSLPAI 758
               :|::|  |
Db        356 AERQALDFI 364


 Search completed: May  5, 2004, 11:09:57
 Job time : 50 secs
```